Research & Development

2026

Mechanical Engineering Letters 2026

Technical-Scientific Journal supported by the Institute of Technology, Hungarian University of Agriculture and Life Science (MATE), Gödöllő, published by GATE Nonprofit Ltd.

HU ISSN 2060-3789

Páter K. u. 1., Gödöllő, H-2103 Hungary

Volume 27 (2026)

Contents

Ebrahim Mirzaiee ASRAMI, Viktor MEDINA, Miklós DARÓCZI:
A multi-stage framework for cost-optimal predictive maintenance: From cost-sensitive classification to dynamic survival modeling:
A comparative analysis using the Scania component X dataset 5

Liman HARTAWAN, Dani RUSIRAWAN, Istvan FARKAS:
Design and operational analysis of photovoltaic-powered IoT systems for smart farming 16

Borisz VÁRKONYI, Viktor ERDÉLYI,
Development of a smart reflex measuring fencing target 30

László ZSIDAI, József NAGY, Attila LÁGYMÁNYOSI:
Solutions for digitizing the human foot orthopedic purposes (brief overview) 42

László KÁTAI, Péter GÁRDONYI, István NAGY:
Motion analysis of a 3D ankle model 67

Márk László LENKEI, Zsigmond Gábor SZALAY, István SZABÓ:
Exploring artificial intelligence LLM adaptation strategies
a study of Retrieval-Augmented Generation (RAG) and Low-Rank Adaptation (LoRA) 81

Abraham Kiprop JUMA, Zoltán KURJÁK, Janos BEKE:
Numerical comparison of flat plate absorber configurations for solar air heaters using ansys fluent 92

Benjámin ALEXE, Zoltán KÁROLY, Gábor KALÁCSKA:
Abrasive damage of mechanical face seals in Martian regolith 103

Tamin DEEB, László KÁTAI:
Vibration Sensors: principles, performance, and applications —
a literature review 118

A multi-stage framework for cost-optimal predictive maintenance: From cost-sensitive classification to dynamic survival modeling: A comparative analysis using the Scania component X dataset

Ebrahim Mirzaiee ASRAMI[1], Viktor MEDINA[2], Miklós DARÓCZI[2]

[1] *Doctoral Program of Mechanical Engineering, MATE, Gödöllő,*
Hungarian University of Agriculture and Life Sciences,
MATE, Gödöllő

[2] *Department of Engineering Management, Institute of Technology,*
Hungarian University of Agriculture and Life Sciences,
MATE, Gödöllő

Abstract

This study presents a comprehensive comparative analysis of maintenance strategies for heavy-duty trucks using the SCANIA Component X dataset. We implement and evaluate multiple machine learning approaches, including Logistic Regression, Multi-Layer Perceptron (MLP) Neural Networks, LightGBM, Random Forest, and survival analysis methods (Cox Proportional Hazards and Random Survival Forest). Our analysis focuses on cost-sensitive decision-making, recognizing the asymmetric costs of false negatives (missed failures) versus false positives (unnecessary maintenance). The results demonstrate that cost-sensitive approaches achieve up to 12.2% cost reduction compared to baseline strategies, with the Economic threshold-based strategy achieving an optimal balance between maintenance coverage and cost efficiency. The Random Forest model achieved 96.7% accuracy, while LightGBM achieved the lowest total cost at $52,364 among classification models. Survival analysis models showed strong discriminative ability, with Random Survival Forest achieving a concordance index of 0.895.

Keywords

predictive maintenance, machine learning, cost-sensitive learning, SCANIA, heavy-duty trucks, Random Forest, survival analysis

1. Introduction

Predictive maintenance (PdM) has emerged as a transformative approach in the automotive industry, leveraging advanced analytics and machine learning to anticipate component failures before they occur (Kharazian et al., 2024). Unlike

traditional reactive maintenance strategies that address failures after they happen, or preventive maintenance that follows fixed schedules regardless of actual component condition, predictive maintenance uses real-time operational data to optimize maintenance timing and resource allocation.

The economic implications of maintenance strategy selection are substantial. In the heavy-duty trucking industry, unexpected breakdowns can result in significant costs including emergency repairs, towing expenses, cargo delays, and potential safety hazards (Moat and Coleman, 2021). Conversely, excessive preventive maintenance leads to unnecessary part replacements and vehicle downtime. This creates a fundamental trade-off that cost-sensitive machine learning approaches aim to optimize.

The SCANIA Component X dataset represents a significant contribution to the predictive maintenance research community, providing real-world multivariate time series data from an internationally recognized truck manufacturer (Kharazian et al., 2024). This dataset addresses a critical gap in the field, as Original Equipment Manufacturers (OEMs) typically maintain strict confidentiality over operational data.

This study makes several contributions: First, we implement and compare multiple classification algorithms evaluating their performance using both traditional accuracy metrics and cost-based evaluation frameworks. Second, we develop and analyze cost-sensitive decision policies that explicitly account for the asymmetric costs of different error types. Third, we provide practical recommendations for maintenance strategy selection.

2. Related work

2.1 Predictive Maintenance in Heavy-Duty Vehicles

The application of machine learning to predictive maintenance has grown substantially in recent years. Ranasinghe et al. (2019) developed prognostic methodologies for scenarios with limited failure data availability. The SCANIA Air Pressure System (APS) dataset has served as an important benchmark (UCI Machine Learning Repository, 2017). Numerous studies have leveraged this dataset including work on handling missing values (Rafsunjani et al., 2019), cost-sensitive classification (Syed et al., 2020), and deep learning approaches (Taghandiki and DallakehNejad, 2023).

2.2 Cost-Sensitive Learning

Traditional classification metrics such as accuracy can be misleading in maintenance contexts where the costs of different error types vary substantially (Beikmohammadi et al., 2024). False negatives typically incur much higher costs than false positives. Akarte and Hemachandra (2018) applied boosting methods with cost-sensitive modifications. Beikmohammadi et al. (2024) developed a

cost-sensitive transformer model specifically designed for prognostics under highly imbalanced industrial data conditions.

2. Dataset description

3.1 SCANIA Component X Dataset Overview

The SCANIA Component X dataset comprises operational data collected from heavy-duty trucks, including multivariate time series readouts, repair records, and vehicle specifications (Kharazian et al., 2024). The training set contains 1,122,452 observations from 23,550 unique vehicles with 107 columns. The operational data includes 14 anonymized variables organized as numerical counters and multi-dimensional histograms.

3.2 Class Structure and Imbalance

The classification task involves predicting urgency level: Class 0 (>48 time steps before failure), Class 1 (24-48), Class 2 (12-24), Class 3 (6-12), and Class 4 (0-6, imminent failure). As shown in Figure 1, the dataset exhibits significant class imbalance with approximately 97% of observations belonging to Class 0 (No Issue). This imbalance reflects real-world conditions where most components operate normally at any given time, but poses significant challenges for machine learning models attempting to detect the minority failure classes.

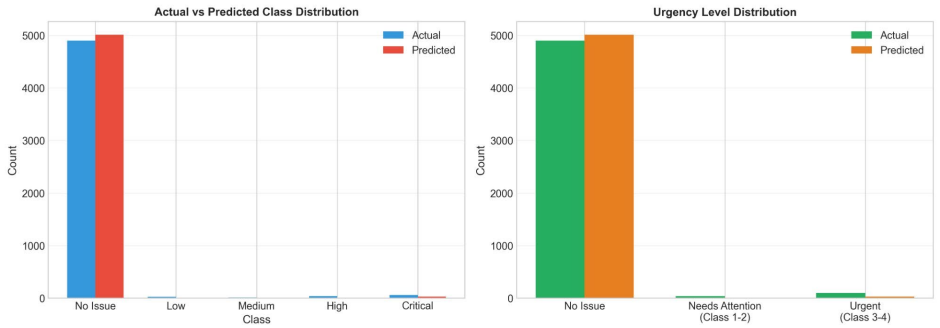


Figure 1. Actual vs Predicted Class Distribution showing severe class imbalance with ~97% Class 0 (No Issue). The right panel groups classes into urgency levels.

3.3 Cost Matrix

A key contribution of the SCANIA dataset is the provision of an expert-defined cost matrix that quantifies the economic impact of different prediction outcomes. The cost of missing an imminent failure (Class 4 predicted as Class 0) is $500, representing potential breakdown costs, emergency repairs, and logistical disruptions. In contrast, a false positive (Class 0 predicted as Class 4) costs only $10, representing the cost of an unnecessary workshop inspection. This 50:1 cost ratio fundamentally shapes optimal decision strategies.

Table 1. Cost matrix for prediction outcomes (USD)

	PRED 0	PRED 1	PRED 2	PRED 3	PRED 4
ACTUAL 0	0	7	8	9	10
ACTUAL 1	200	0	7	8	9
ACTUAL 2	300	200	0	7	8
ACTUAL 3	400	300	200	0	7
ACTUAL 4	500	400	300	200	0

4. Methology

4.1 Classification Models

We implemented four classification algorithms: (1) Logistic Regression - a linear classifier providing interpretable coefficients; (2) Multi-Layer Perceptron (MLP) Neural Network with two hidden layers and dropout regularization; (3) LightGBM - a gradient boosting framework using histogram-based algorithms; and (4) Random Forest - an ensemble method providing feature importance rankings.

4.2 Survival Analysis Models

We implemented Cox Proportional Hazards and Random Survival Forest (RSF) models. These methods model time-to-failure directly, enabling nuanced risk assessment while properly handling censored observations. Figure 2 presents the survival analysis results, including mean survival curves, risk score distributions, model performance comparison (C-index), and feature importance for RSF.

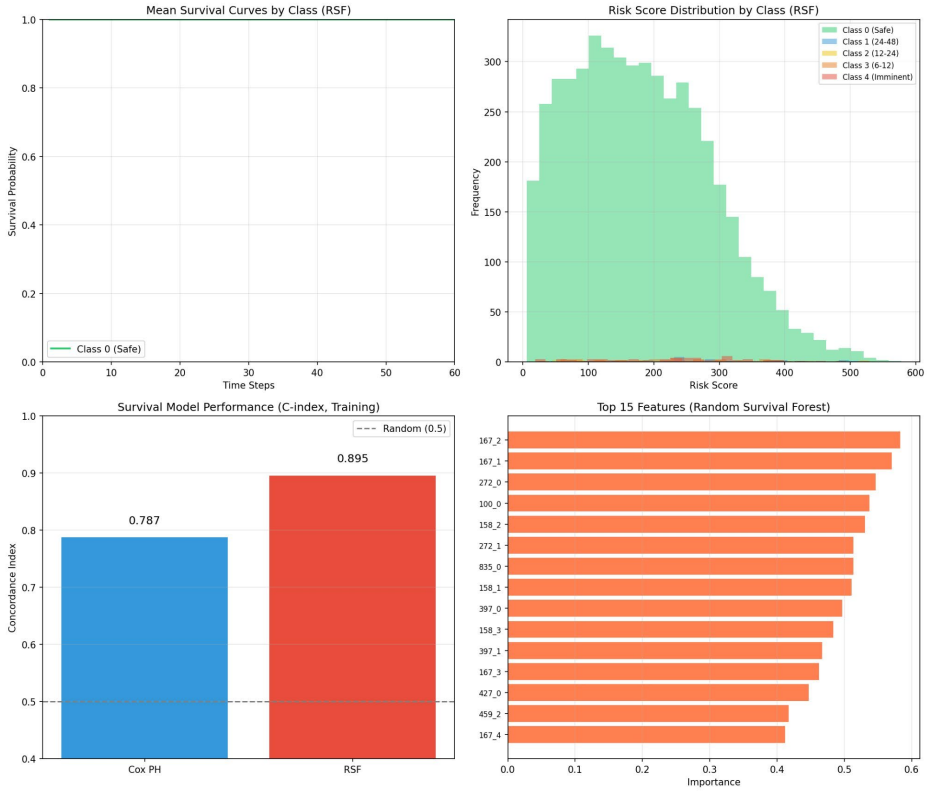


Figure 2. Survival analysis results: (a) Mean survival curves by class showing probability of component survival over time; (b) Risk score distribution by urgency class; (c) C-index comparison showing RSF (0.895) outperforming Cox PH (0.787); (d) Top 15 features for Random Survival Forest.

4.3 Maintenance Strategies

We evaluated four strategies: (1) Do Nothing - baseline with no predictive maintenance; (2) Model Prediction - direct use of predicted class labels; (3) Maintain All - conservative strategy scheduling maintenance for all vehicles; and (4) Cost-Sensitive (Economic) - threshold-based approach optimizing expected cost by considering prediction uncertainties and asymmetric costs.

5. Results

5.1 Model Performance Comparison

Figure 3 presents a comprehensive comparison of all models evaluated. A striking finding is the disconnect between accuracy and cost performance. As

shown in the accuracy vs cost trade-off plot (top right), MLP Neural Network and Random Forest achieve the highest accuracy (~97%), yet incur higher total costs than LightGBM (90.2% accuracy). This paradox occurs because high-accuracy models tend to predict the majority Class 0, missing critical failure cases that carry high misclassification costs.

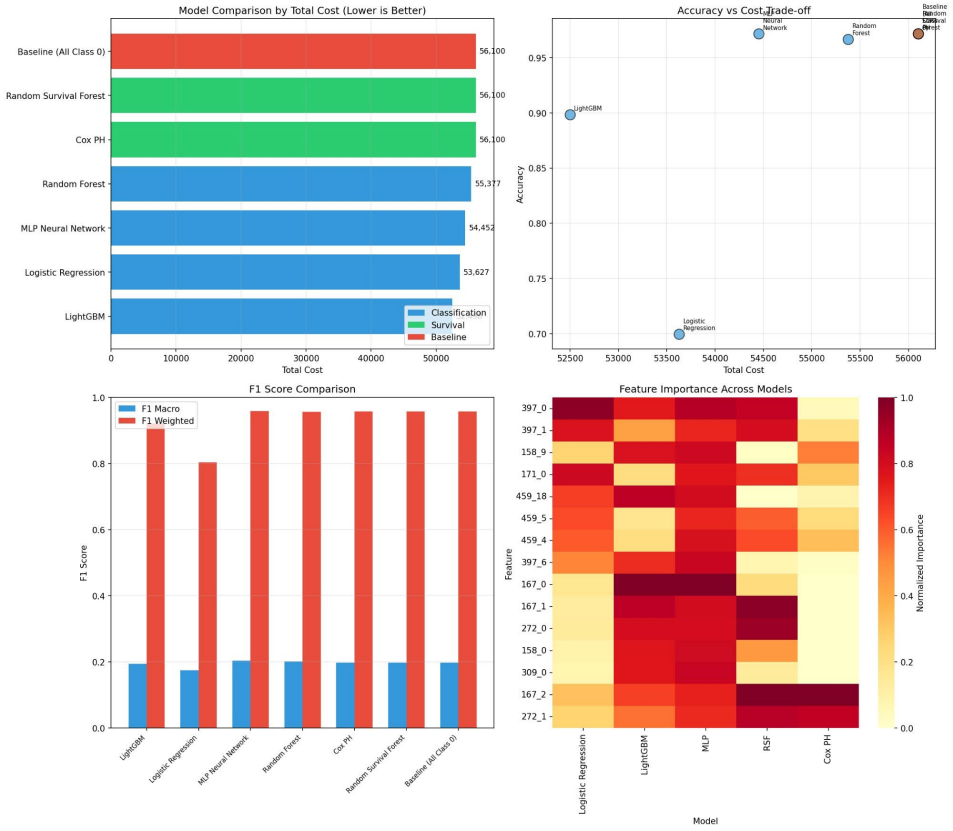


Figure 3. Comprehensive model comparison: (a) Total cost comparison showing LightGBM achieves lowest cost; (b) Accuracy vs Cost trade-off revealing the paradox where highest accuracy models incur higher costs; (c) F1 Score comparison; (d) Feature importance heatmap across models showing consistent patterns.

Table 2. Classification model performance comparison

MODEL	ACCURACY	TOTAL COST	F1 WEIGHTED
LOGISTIC REGRESSION	70.0%	$53,627	0.80
MLP NEURAL NETWORK	97.3%	$54,452	0.97
LIGHTGBM	90.2%	$52,364	0.94
RANDOM FOREST	96.7%	$55,377	0.97

5.2 Confusion Matrix and Cost Analysis

Figure 4 displays the confusion matrix alongside the associated cost distribution. The left panel shows that the model correctly classifies 4,876 Class 0 instances but misclassifies most failure cases (Classes 1-4) as Class 0. The right panel translates these misclassifications into costs, revealing that missed Class 4 failures (59 instances × $500 = $29,500) contribute the largest portion of total cost, followed by missed Class 3 failures ($16,400).

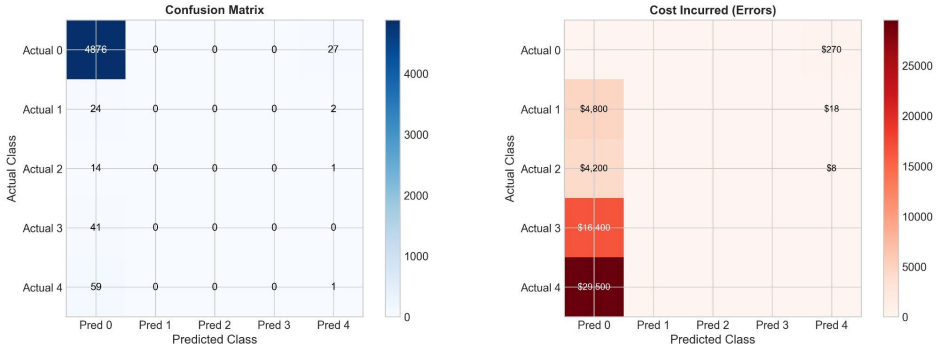


Figure 4. Confusion matrix (left) and associated cost incurred by errors (right). The cost heatmap shows that false negatives for Classes 3-4 (missed urgent failures) dominate total costs despite being numerically rare.

5.3 Maintenance Strategy Comparison

Figure 5 compares the four maintenance strategies on the test set of 5,045 vehicles. The Cost-Sensitive (Economic) strategy achieves the lowest total cost ($49,239), representing a 12.2% reduction compared to the Do Nothing baseline ($56,100). Notably, the simple Maintain All strategy achieves $49,671,

capturing 94% of the optimal savings with significantly less implementation complexity.

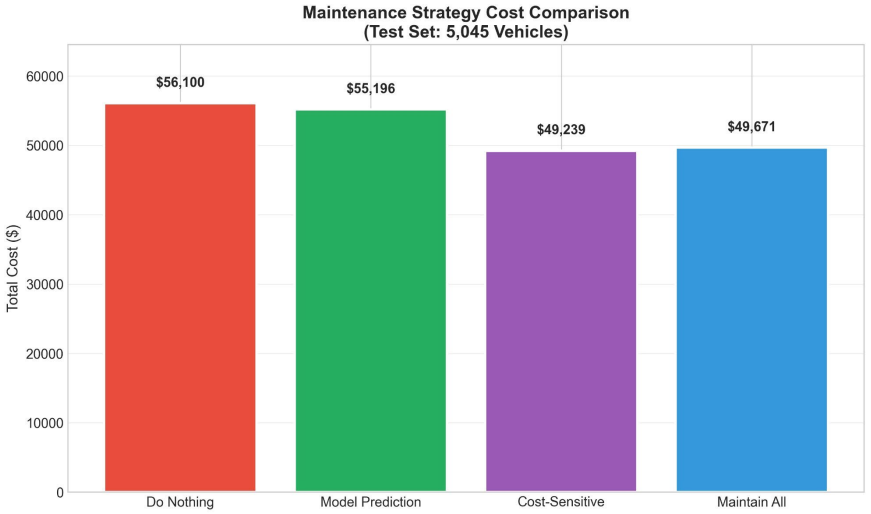


Figure 5. Maintenance strategy cost comparison on test set (5,045 vehicles). Cost-Sensitive strategy achieves optimal $49,239, while Maintain All achieves nearly equivalent $49,671.

Table 3. Maintenance strategy comparison

STRATEGY	TOTAL COST	SAVINGS VS BASELINE
DO NOTHING (BASELINE)	$56,100	-
MODEL PREDICTION	$55,196	1.6%
MAINTAIN ALL	$49,671	11.5%
COST-SENSITIVE (ECONOMIC)	$49,239	12.2%

5.4 Decision Policy Analysis

Figure 6 provides detailed analysis of the decision policy. The top-left panel shows that 5,014 vehicles require no action while only 31 are classified as critical. The bottom-left panel displays the distribution of net maintenance benefit, with the decision threshold (red dashed line) separating vehicles where maintenance provides positive expected value. The bottom-right confusion matrix confirms the model's strong performance on Class 0 but limited ability to detect minority failure classes.

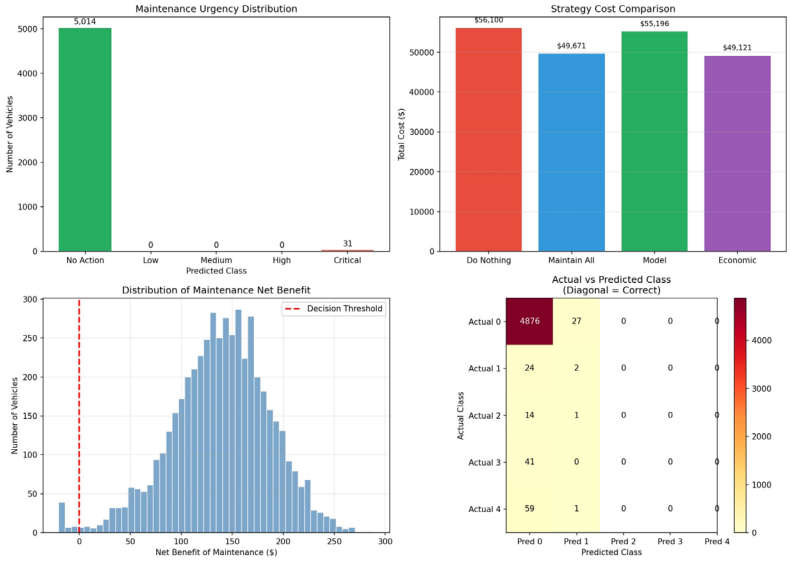


Figure 6. Decision policy analysis: (a) Maintenance urgency distribution; (b) Strategy cost comparison; (c) Net benefit distribution with decision threshold; (d) Detailed confusion matrix.

5.5 Feature Importance Analysis

Figure 7 presents the top 20 most important features from the Random Forest model. Features from histogram variable "167" (bins 1, 0, and 6) emerge as the most predictive, followed by multiple bins from variable "397". This pattern is consistent across different model architectures (as shown in Figure 3d), validating the robustness of these operational parameters as failure indicators.

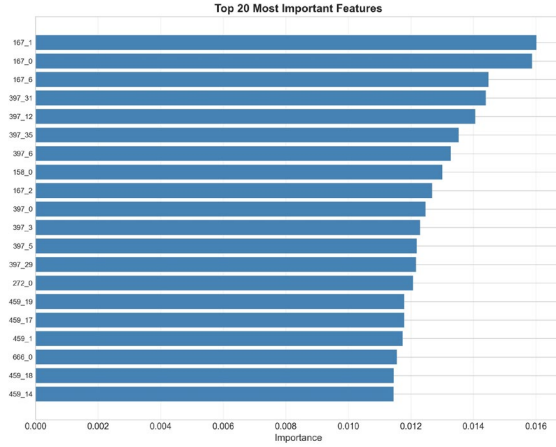


Figure 7. Top 20 most important features from Random Forest. Variable "167" bins dominate, followed by variable "397" bins, suggesting these operational parameters are key failure predictors.

6. Discussion

Our findings highlight several important insights for predictive maintenance practice:

First, the accuracy-cost disconnect demonstrates that standard machine learning metrics can be misleading in industrial applications where error costs are asymmetric. As shown in Figure 3, models optimized for accuracy (MLP, Random Forest at ~97%) perform worse on cost metrics than LightGBM at 90.2% accuracy. This occurs because accuracy-optimized models favor the majority class, missing critical failures.

Second, the strong performance of the simple Maintain All strategy (capturing 94% of optimal savings) suggests that in scenarios with highly asymmetric costs (50:1 ratio), conservative approaches may be near-optimal. The incremental benefit of sophisticated cost-sensitive optimization (0.7% additional savings) must be weighed against implementation complexity.

Third, feature importance analysis reveals consistent patterns across models, with histogram variables "167" and "397" emerging as dominant predictors. This convergence validates the robustness of the identified predictive signals and provides actionable insights for maintenance diagnostics.

Limitations include: the cost matrix based on expert estimates may vary across contexts; the dataset represents a specific subset of SCANIA trucks; and variable anonymization limits domain knowledge incorporation.

Conclusions

This study presented a comprehensive analysis of cost-sensitive machine learning approaches for predictive maintenance using the SCANIA Component X dataset. Key findings include:

(1) Traditional accuracy metrics can be misleading - LightGBM achieved lowest cost ($52,364) despite not having highest accuracy (90.2% vs 97.3% for MLP).

(2) Cost-sensitive strategies achieve up to 12.2% cost reduction compared to baseline, with the Economic strategy achieving optimal $49,239.

(3) Simple conservative strategies (Maintain All at $49,671) capture 94% of optimal savings, offering practical value without sophisticated optimization.

(4) Random Survival Forest achieves strong discriminative ability (C-index 0.895), substantially outperforming Cox Proportional Hazards (0.787).

Future work should integrate survival analysis predictions into cost-sensitive decision frameworks, explore online learning approaches to handle distribution drift, and investigate additional data sources such as telematics, environmental factors, and driver behavior patterns.

Acknowledgements

The author thanks SCANIA CV AB and Halmstad University researchers for making the Component X dataset publicly available, and Tamim Deeb and Sahab Alkhouli for their support in not giving up.

References

[1] Abidi, M.H., Mohammed, M.K., and Alkhalefah, H. (2020) Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing, Sustainability, 14(6), 3387.

[2] Akarte, M.M. and Hemachandra, N. (2018) Boosting Based Cost Sensitive Approach for APS Failure Prediction, IEEE IEEM, pp. 1-5.

[3] Beikmohammadi, A., Faez, K., and Motallebi, H. (2024) Cost-Sensitive Transformer for Remaining Useful Life Prediction, IEEE Trans. Industrial Informatics, 20(3), 4567-4578.

[4]Ke, G., Meng, Q., Finley, T., et al. (2022) LightGBM: A Highly Efficient Gradient Boosting Decision Tree, NeurIPS, 30, 3146-3154.

[5]Kharazian, Z., Lindgren, T., Magnusson, M., and Bouguelia, M.R. (2024) SCANIA Component X Dataset: A Real-World Multivariate Time Series Dataset for Predictive Maintenance, arXiv:2401.15199.

[6]Moat, A. and Coleman, S. (2021) Industrial Analytics for Predictive Maintenance of Commercial Vehicles, Applied Sciences, 11(21), 10336.

[7]Rafsunjani, S., et al. (2019) An Empirical Comparison of Missing Value Imputation Techniques on APS Failure Prediction, IJITCS, 11(2), 21-29.

[8]Ranasinghe, R.H., et al. (2019) Machine Learning Approaches for Predictive Maintenance: A Case Study, IEEE ICIIS, pp. 1-6.

[9]Syed, M.N., Khan, S., and Ahmad, A. (2020) Cost-Sensitive Classification for Air Pressure System Failure Prediction, Computers and Industrial Engineering, 150, 106856.

[10]Taghandiki, M. and DallakehNejad, A. (2023) Deep Learning Approaches for Predictive Maintenance of Heavy-Duty Vehicles, Expert Systems with Applications, 213, 118882.

[11]UCI Machine Learning Repository (2017) APS Failure at Scania Trucks Data Set.

Design and operational analysis of photovoltaic-powered IoT systems for smart farming

Liman HARTAWAN[1,2], Dani RUSIRAWAN[2], Istvan FARKAS[3]

[1] *Doctoral Program of Mechanical Engineering, Hungarian University of Agriculture and Life Sciences, MATE, Gödöllő, Hungary*

[2]*Department of Mechanical Engineering, Institut Teknologi Nasional Bandung, Bandung, Indonesia*

[3]*Institute of Technology, Hungarian University of Agriculture and Life Sciences, MATE, Gödöllő, Hungary*

Abstract

The photovoltaic (PV) as energy harvester to powered Internet of Things technologies for smart farming in arable land with limited resources or no connection to the electricity grid is increasingly promising. This study describes a preliminary design and energy performance estimates solar-powered Internet of Things (IoT) technology for smart farming. The energy consumption requirements of smart farming for the components used and the potential power output from the PV are presented to investigate the energy sufficiency. Based on study, it has been found that for the specified system, the energy consumption for the sensor nodes in smart farming system for 24 hours is 2.52 Wh, the minimum PV capacity to support the operation of the sensor node during the charging process is 0.64 W, and the minimum battery capacity for 12 hours of operation before recharging is 410 mAh. This study provides initial design parameters for the implementation of a small-scale smart farming system that includes sensor nodes powered by the PV for arable lands, which will be investigated experimentally in further research.

Keywords

energy harvester, arable land, energy sufficiency, energy consumption, design parameters

1. Introduction

Smart farming technology now is widely use to increase agricultural productivity while reducing adverse impacts on the environment. At the same time, the smart farming technologies are emerging as important solutions to address the challenges of food security, climate change, and resource scarcity (Raj & Prahadeeswaran, 2025). In its application, smart farming uses sensors installed spread across agricultural land. In general, smart farming on

agricultural land uses rechargeable batteries which still difficult to maintain operate continuously. The use of the PV as an energy harvester for charging the battery in smart farming is increasingly promising, which correlated with lower microcontroller power consumption and increased the PV efficiency (Luo et al., 2024). This is especially in environments with limited resources or no connection to the electricity grid, and in tropical areas which always exposed to high intensity sunlight throughout the year. Planting area divide into indoor and outdoor scenarios, with Crop beds, Greenhouse and Hydroponic environments for indoor scenario, and Arable Lands and Orchard environments for outdoor scenarios (Navarro et al., 2020), as shown in Fig. 1. This study focuses on the Arable Lands environments as areas without electricity sources and always exposed to sunlight.

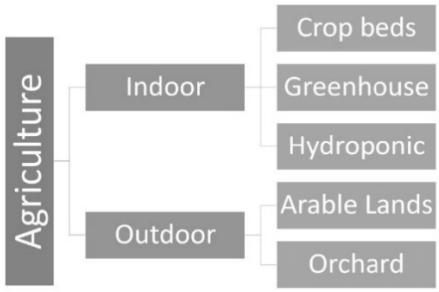


Figure 1. Typical scenarios and environments in agriculture (Navarro et al., 2020)

Indonesia is a country located on the equator which is a tropical area, so it receives sunlight throughout the year with characteristic of irradiation 4.8 kWh/m^2/day (Silalahi et al., 2021). With low power sensors for smart farming parameter and increased the PV efficiency, PV size can be smaller in these areas. Its small size can be applied for portable smart farming equipment, which making easier for farmers to install and more reasonable as energy harvester.
This study provides preliminary design and energy performance estimates the PV-powered Internet of Things technologies for smart farming, which will be investigated experimentally in further research. The energy consumption requirements of smart farming for the components used and the potential power output from photovoltaic are presented to investigate the energy sufficiency. Hereafter, the material and method, results and discussion, and conclusions, also presented in this paper.

1.1 Overview of smart farming

Smart farming is a farm management concept that employs techniques and technologies such as sensors to collect data correlated to agricultural production, communication networks to interconnecting devices, and management information systems to manage and analysed data that helps boosting

productivity and minimizing waste (Navarro et al., 2020). An illustrated IoT based smart farming, is shown in Fig. 2.

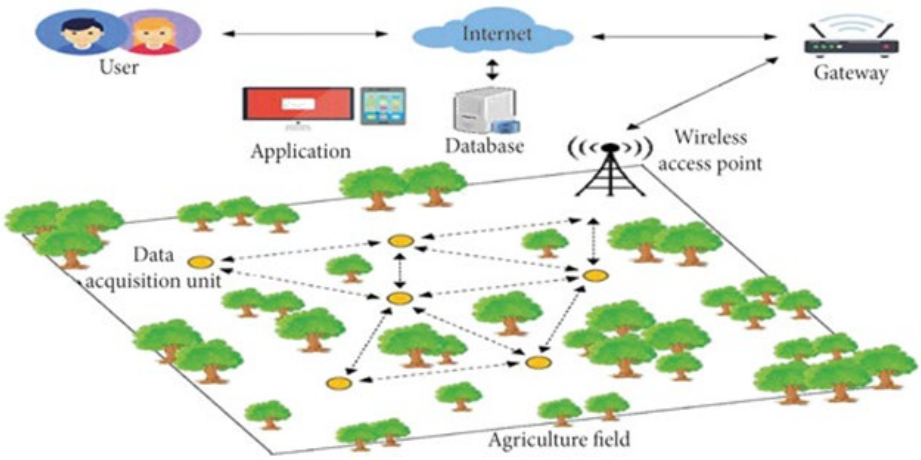


Figure 2. Illustration of smart farming with IoT based system (Li et al., 2020)

In agriculture field, sensor nodes that collect environmental data such as soil moisture, air temperature and humidity, light intensity, and CO_2 levels, are integrated by a communication network and then transmitted to the internet using Internet of Things (IoT) technology for processing and analysis. The communication network between sensor nodes to gateway can be wireless or a combination of wireless and wired, which commonly used is wireless called Wireless Sensors Network (WSN) (Musa et al., 2024), (Villa-Henriksen et al., 2020). The IoT architecture is divided into three layers as device, network and application layers, as shown in Fig. 3.

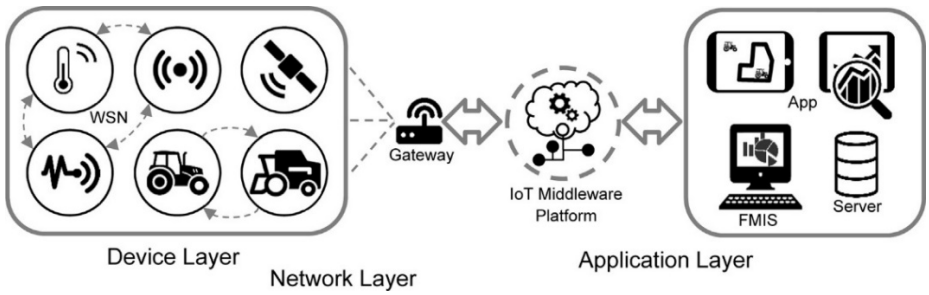
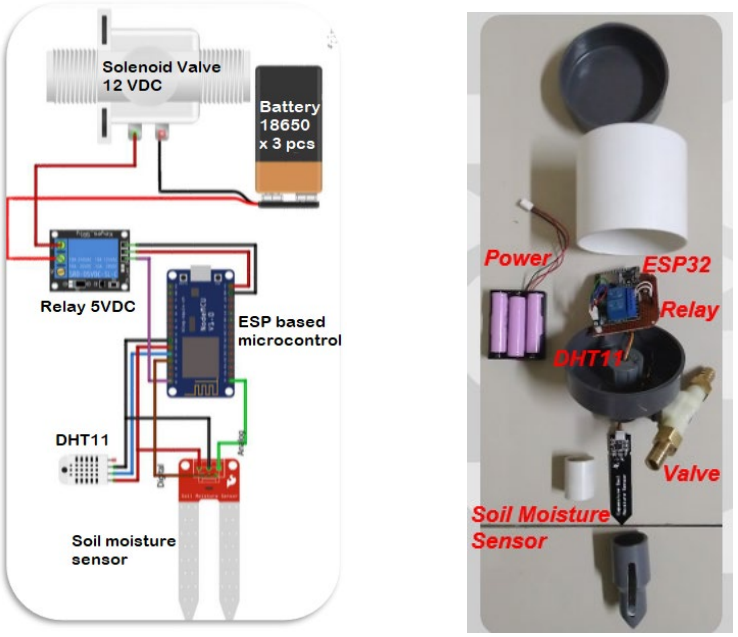


Figure 3. IoT Architecture (Villa-Henriksen et al., 2020)

The device layer contains sensors to collect data or actuators to perform required actions, which can be connected to each other and are called as WSN (Wireless Sensor Network) for sensors. At the network layer, data is transmitted from the device layer to the internet through a gateway. Then in the application layer, data is collected and processed for analysis or visualization.

1.2. Previous studies

Sensor nodes require power not only for the sensing process, but the data transmission process also consumes power. The use of photovoltaics as energy harvesters is an effective method for converting solar energy into electricity to recharge batteries in sensor nodes, so that the system can operate continuously. In previous study, sensor node as an automatic sprinkler based on soil moisture has been built, which powered by rechargeable batteries (Hartawan et al., 2022), as shown in Fig. 4. The use of photovoltaic has also been applied to the automatic sprinkler but has not been optimized (Anggraeni et al., 2022), as shown in Fig. 5.



a) b)

Figure 4. Automatic Springkler a) Wiring diagram b) Components with PVC enclosure (Hartawan et al., 2022)

Figure 5. PV-powered Automatic Sprinkler a) Wiring diagram b) Realization (Anggraeni et al., 2022)

Moreover, by selecting low-power sensors, enabling sleep-mode on the microcontroller, choosing the suitable communication protocol and setting how data is transmitted, will reduce power consumption on sensor nodes. The ESP-NOW communication protocol was chosen in this analysis because it is associated with the use of ESP-based microcontrollers due to its low cost and embedded technology.

2. System design and performance evaluation

In this section the assessment of solar radiation and the details of the designed system configuration will be discussed.

2.1. Assessment of solar energy in Bandung (Indonesia)

The average solar radiation in Bandung, Indonesia on July 10, 2018 was 445.8 W/m^2 (Setyawan & Sutandi, 2020), as shown in Fig. 6. This value for 12 hours under clear sky condition is equivalent to 5.33 kWh/m^2 (Setyawan & Sutandi, 2020). Solar radiation in Bandung ranges from 4.57 to 5.35 kWh/m^2, with the lowest value in November and the highest value in July (Rumbayan et al., 2012). The monthly average solar irradiation data of 22 years period from NASA website (http://eosweb.larc.nasa.gov) for Bandung city, as shown in Table 1.

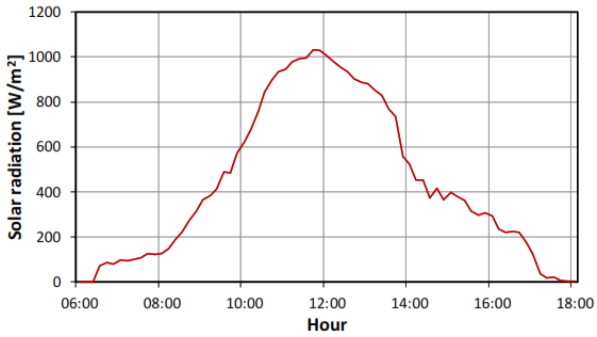


Figure 6. Solar radiation measured, July 10, 2018 (Setyawan & Sutandi, 2020)

Table. 1. The monthly solar irradiation data (in kWh/m2) of Bandung (Rumbayan et al., 2012)

City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Bandung	4.57	4.75	4.87	4.95	5.02	4.97	5.17	5.35	5.11	4.77	4.7	4.96

The sun hours in Bandung city in year 2025 from website (www.gaisma.com) is 12 hours 4 minutes. In December, 15th the sunrise at 05:32 and the sunset at 18:04 (12 hours 32 minutes), in March, 15th the sunrise at 06:05 and the sunset at 17:58 (11 hours 53 minutes), and in June, 15th the sunrise at 06:00 and the sunset at 17:43 (11 hours 47 minutes), as shown in Fig. 7.

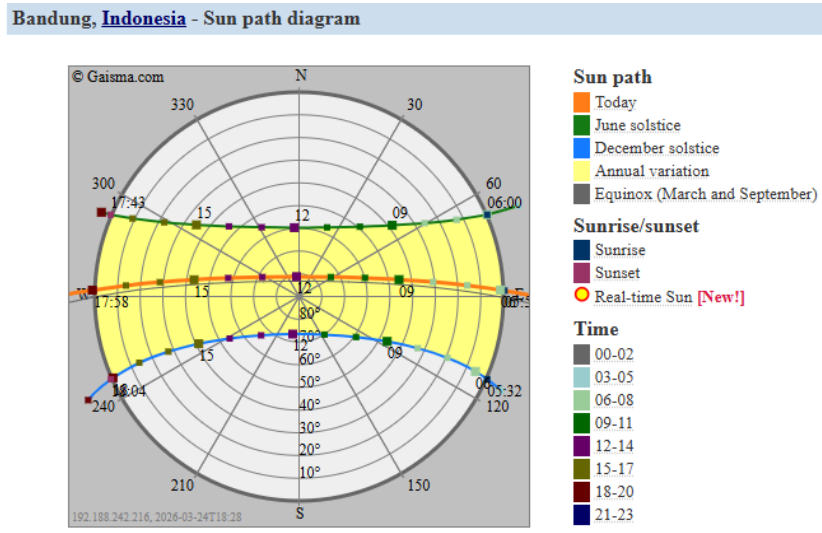
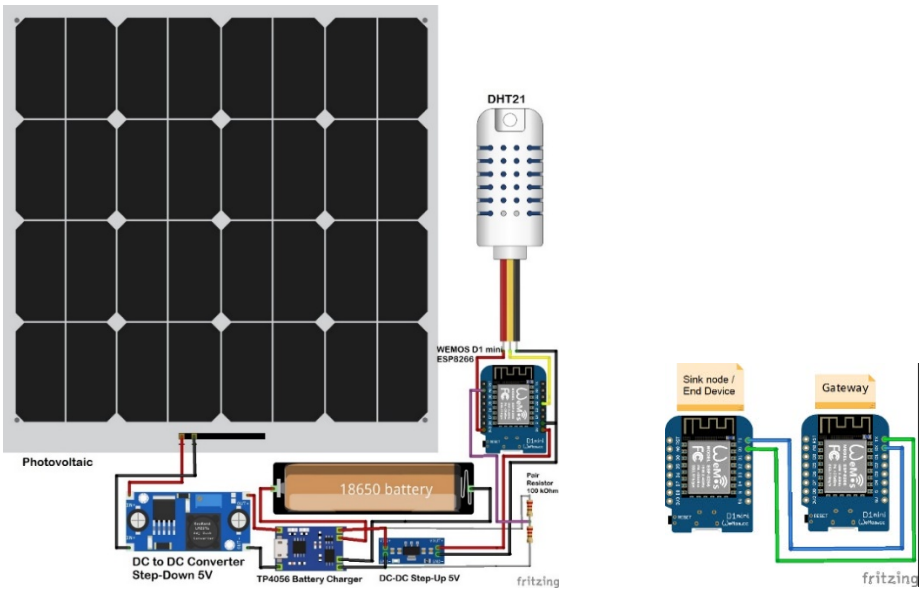


Figure 7. Sun path diagram of Bandung city in Indonesia in 2025 (Tukiainen, 2026)

2.2. System configuration

The proposed smart farming system integrated sensor nodes, sink node or end device, and gateway, is shown in Fig. 8. Each sensor node contains an air and humidity sensor, an ESP-based microcontroller, a battery, a battery charger, a PV module, a DC Step-Down, a DC Step-Up, and a pair of resistors to detect the battery voltage (Hartawan et al., 2024), as shown in Fig. 8a). The PV are used as energy harvester to recharging the battery in each sensor node, the system namely Solar Energy Harvesting – Wireless Sensors Network (SEH-WSN) (Sharma et al., 2019).



a)

b)

Figure 8. Wiring diagram of a) sensor nodes, b) Sink node & Gateway (Hartawan et al., 2024)

Sensor nodes are connected to other nodes to transmit monitored data using Wireless Sensors Network (WSN) from the farthest node to the nearest sink node or end device, and is shown in Fig. 9. From the sink node or end device, the monitoring data is sent to the internet through the gateway.

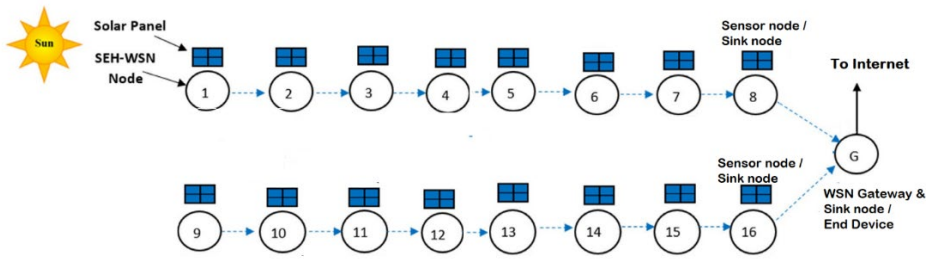


Figure 9. Framework for smart farming monitoring using SEH-WSN, modified from (Sharma et al., 2019)

2.3. The photovoltaic (PV) selection

A monocrystalline type of PV module was carefully chosen for this study due to their higher efficiency for power generation and technically and commercially feasible today and will remain so in the future (Altaye et al., 2025). With higher PV efficiency, the size and weight will be smaller, thus reducing the size of the sensor node. Based on data from (National Renewable Energy Laboratory (NREL), 2025) the cell efficiency for the single crystal (non-concentrator) type is 26.6%, as shown in Fig. 10.

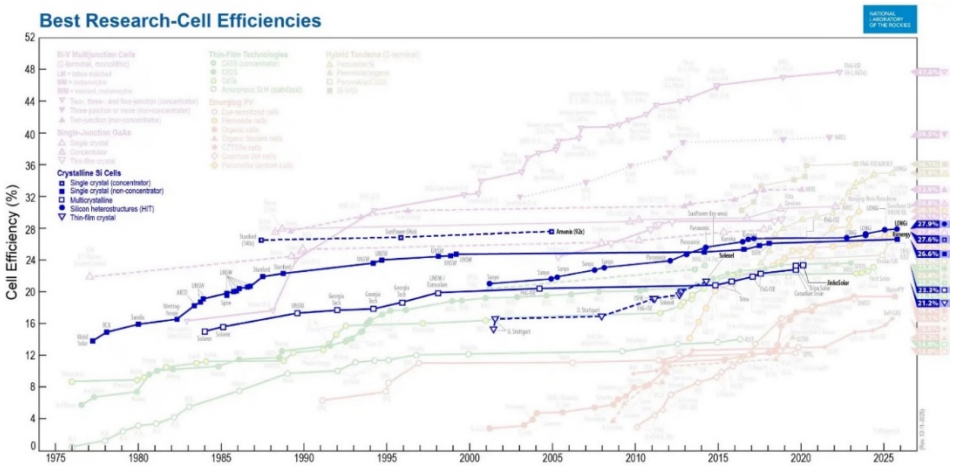


Figure 10. The timeline chart of the highest research cell efficiencies for Cryctalline Si Cells photovoltaic technologies from 1975 to 2025 (NREL, 2025)

2.4. Communication protocol selection

ESP-NOW is wireless communication protocol developed by Espressif Systems, a semiconductor company, for two or more ESP based microcontrollers to

communicate without using internet access. The ESP-NOW protocol uses the IEEE 802.11 standard but reduce the upper five layers to one layer, which can reduce the power consumption and increases the range and reliability of the connection, as shown in Fig. 11.

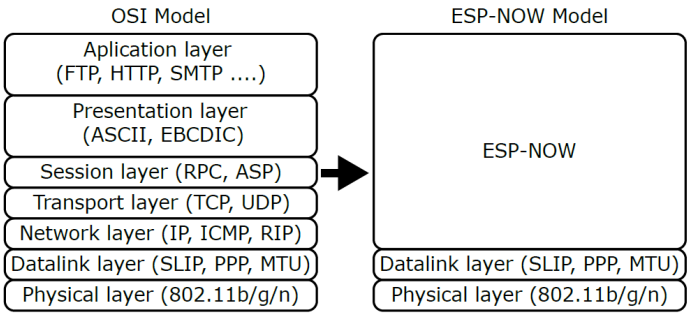


Figure 11. Comparison of the network layer model of standard Wi-Fi connection (OSI) and ESP-NOW (Sokol et al., 2024)

Functional testing of the ESP-NOW protocol using a WEMOS D1 mini (ESP8266-based microcontroller) inside a PVC case showed that the transmission range reached 45 meters (Hartawan et al., 2025). Due to the limited sensing area of the sensor, the transmission range of ESP-NOW is still suitable for use as a Wireless Sensor Network (WSN) for smart agriculture. To cover large areas in agricultural land, sensor nodes are connected using tree topology from the farthest node to the nearest sink node or end device.

2.5. Performance prediction

The performance of the systems is evaluated from the energy consumption for sensor node by obtaining total current hours of the components, which can be expressed using Eq. 1:

$$I_{Avq} = \frac{(I_r \, x \, T_r)}{T_c} + \frac{(I_i \, x \, T_i)}{T_r} \, mA$$

where:

I_{Avg} = Average current consumption
I_r = Run time current consumption
T_r = Running time in seconds
I_r = Idle time current consumption
T_i = Idle time in seconds
T_c = Cycle time in seconds (which Active time + Light sleep = 60 second)

For sensing voltage, the flow current is computed by Eq. 2:

$$I_{sens} = \frac{V_{Batt}}{R_{total}}$$

where:

R_{total} = R_1 + R_2 (Ohm).

Because of using DC Step-Up, the battery current must consider the efficiency, which can be expressed using Eq. 3:

$$I_{Batt} = \frac{I_{total} \, x \, V_{output}}{V_{Batt} \, x \, \eta_{Batt}}$$

where:

V_{output} is the operating voltage of the system (V),
V_{Batt} is the battery voltage (V),
η_{Batt} is 85%.

To calculate battery capacity is computed by Eq. 4:

$$I_{Batt} \, x \, T_r = C \; mAh$$

For battery operating time can be expressed using Eq. 5:

$$T_{Op} = \frac{C}{I_{Batt}}$$

2.6. Operating parameters

The proposed sensor nodes in smart farming system were analysed using parameter H_{sun} is 12 hours represents the number of hours of sunshine per day (h) in Indonesia especially in Bandung. The $\eta_{PV\ module}$ is 26.6% represents the efficiency of the monocrystalline type photovoltaic, $\eta_{charge\ controller}$ is 85% represents the charge controller efficiency, and η_{DOD} is presume 80% represents the depth of discharge of Lithium-Ion battery. The detailed input parameters for this analysis were summarized in Table 2.

Table 2. Basic operating parameters

Parameters	Specification
Photovoltaic type	Monocrystalin ($\eta_{PV\ modulel}$ = 26.6%)
Battery type	Lithium-Ion (η_{DOD} = 80%)
Communication protocol	ESP-NOW
Microcontroller	WEMOS D1 mini (ESP8266); 120 mA (active) 1.35 mA (light sleep)
Air Humidity & Temperature sensor	DHT21; 1.5 mA (active) 0.3 mA (idle/sleep)
Charge controller	TP4056 ($\eta_{charge\ controller}$ = 85%)
A pair of resistor	100 kOhm

3. Results and discussion

The sensor node for this study contains an air and humidity sensor, an ESP-based microcontroller, a battery charger, a DC Step-Down, a DC Step-Up, and a pair of resistors to detect the battery voltage. The power consumption for those components is shown in Table 3.

Table 3. Power consumption of sensor node components

Component	Power consumption
ESP microcontroller	$I_{ESP} = \frac{(I_{Act}\ x\ T_{Act})}{T_c} + \frac{(I_i\ x\ T_i)}{T_c}$ $= \frac{(120\ mA\ x\ 10\ s)}{60\ s} + \frac{(1.35\ x\ 50\ s)}{60\ s}$ = 21.12 mA
Air Humidity & Temperature sensor	T_c = 240 s; T_{act} = 1 s; T_i = 239 s $I_{DHT} = \frac{(I_{Act}\ x\ T_{Act})}{T_c} + \frac{(I_i\ x\ T_i)}{T_c}$; $= \frac{(1.5\ x\ 1)}{240} + \frac{(0.3\ x\ 239)}{240}$ = 0.305 mA
A pair of resistor to detect battery 3.7 V	$R_{total} = R_1 + R_2$ = 100 + 100 kOhm = 200 kOhm
DC Step-Up	$I_{sens} = \frac{V_{Batt}}{R_{total}} = \frac{3.7\ V}{200\ kOhm}$ = 0.0185 mA

The total current is:

$$I_{total} = I_{ESP} + I_{DHT} + I_{Sens}$$

$$I_{total} = 21.12 \text{ mA} + 0.305 \text{ mA} + 0.0185 \text{ mA} = 21.44 \text{ mA} = 0.021 \text{ A}$$

Because of using DC Step-Up, with considering the efficiency (85%), the battery current is:

$$I_{Batt} = \frac{I_{total}\ x\ V_{output}}{V_{Batt}\ x\ \eta_{Batt}} = \frac{21.44\ mA\ x\ 5\ V}{3.7\ V\ x\ 85\%} = 34.08 \text{ mA} = 0.034 \text{ A}$$

The battery capacity to perform sensor nodes operating 24 hours is:

$$I_{Batt}\ x\ T_r = C\ mAh = 34.08 \text{ mA x } 24 \text{ h} = 817.92 \text{ mAh}$$

To find operating time for the battery with capacity 1000 mAh is:

$$T_{Op} = \frac{C}{I_{Batt}} = \frac{1000\ mAh}{34.08\ mA} = 29.34 \text{ h}$$

With a battery capacity of 1000 mAh, the sensor node can operate for 29.34 hours before needing to be recharged. To operate for 12 hours, it can use a battery with minimum capacity of 410 mAh.

The energy for charging the battery is:

$E_{charge} = I_{Batt}$ x V_{output} x H_{Sun} = 0.034 A x 5 V x 12 hours = 0.17 W x 12 hours

= 2.04 Wh = 0.002 kWh

To support the operation of the sensor node during the charging process, the minimum PV capacity is:

$$P_{PV\ module} = \frac{I_{Batt}\ x\ V_{Output}}{\eta_{PV\ module}} = \frac{0.034\ A\ x\ 5\ V}{26.6\ \%} = 0.64 \text{ W}$$

The energy consumption for the sensor node is:

$E_{charge} = I_{total}$ x V_{output} x T_r = 0.021 A x 5 V x 24 hours = 0.105 W x 24 hours

= 2.52 Wh = 0.002 kWh

Conclusions

Overall, this study presented predicted performance and initial design parameters of the photovoltaic-powered Internet of Things technologies for smart farming. The components inside sensor nodes and Wireless Sensor Network (WSN) was presented. The results of the study show that the power consumption of the sensor nodes is low and in locations with long sunshine hours such as in the city of Bandung (Indonesia), the use of photovoltaics as energy harvesters becomes more feasible. The power consumption of the sensor node is 2.52 Wh, with a minimum battery of 410 mAh to support 12 hours of operation before recharging, and a minimum photovoltaic capacity of 0.64 W that can support the charging process during operation.

The result provides a preliminary design of sensor nodes and Wireless Sensors Network (WSN) for smart farming and the power consumption of the systems. This design supports future work prior to the experimental setup focused on developing and testing the system to achieve optimal performance.

Acknowledgements

This work was supported by the Stipendium Hungaricum Programme and by the Doctoral School of Engineering Sciences, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary.

References

[1] Altaye, A., Víg, P., & Farkas, I. (2025). The Assessment of the Potential and Development of Photovoltaic Technology in Energy Production: Review. In *Energy Science and Engineering* (Vol. 13, Number 10, pp. 5030–5050). John Wiley and Sons Ltd. https://doi.org/10.1002/ese3.70179

[2] Anggraeni, N. D., Hartawan, L., Shantika, T., Rusirawan, D., Seres, I., & Farkas, I. (2022). Photovoltaic Utilization as Energy Source For Automatic Sprinklers In Agriculture. In I. Farkas & P. Víg (Eds.), *28th Workshop on Energy and Environment* (p. 21). Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary. https://press.mater.uni-mate.hu/80/2/28th Workshop on Energy and Environment_2022_absztraktkötet.pdf#page=23

[3] Hartawan, L., Rakhmat, G. A., Nugraha, N., Anggraeni, N. D., Maulana, K. B., Ridjali, M. R., Iqbal, M., Mahardika, M. K., Alam, B. S., Fani, J. Al, Amarullah, L. F., Alfadhlih, M. I., Sbastio, M. R., & Purba, A. A. (2024). Design of IoT-based Greenhouse Temperature and Humidity Monitoring System. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *12*(4), 1051. https://doi.org/10.26760/elkomika.v12i4.1051

[4] Hartawan, L., Rusirawan, D., & Farkas, I. (2025). Comparison of Communication for Wireless Sensor Network in Smart Farming Applications. In I. Farkas & P. Víg (Eds.), *31st Workshop on Energy and Environment* (pp. 39–40). Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary. https://press.mater.uni-mate.hu/529/1/31st-Workshop-on-Energy-and-Environment-2025-full.pdf#page=40

[5] Hartawan, L., Shantika, T., Anggraeni, N. D., Rusirawan, D., & Farkas, I. (2022). Smart Farming Application Based On Automatic Sprinkler and Arduino IoT Cloud. In I. Farkas & P. Víg (Eds.), *28th Workshop on Energy and Environment* (p. 35). Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary. https://press.mater.uni-mate.hu/80/2/28th Workshop on Energy and Environment_2022_absztraktkötet.pdf#page=37

[6] Li, W., Awais, M., Ru, W., Shi, W., Ajmal, M., Uddin, S., & Liu, C. (2020). Review of Sensor Network-Based Irrigation Systems Using IoT and Remote Sensing. *Advances in Meteorology*, *2020*. https://doi.org/10.1155/2020/8396164

[7] Luo, K., Chen, Y., Lin, R., Liang, C., & Zhang, Q. (2024). A Portable Agriculture Environmental Sensor with a Photovoltaic Power Supply and Dynamic Active Sleep Scheme. *Electronics (Switzerland)*, *13*(13).

https://doi.org/10.3390/electronics13132606

[8] Musa, P., Sugeru, H., & Wibowo, E. P. (2024). Wireless Sensor Networks for Precision Agriculture: A Review of NPK Sensor Implementations. *Sensors*, *24*(1). https://doi.org/10.3390/s24010051

[9] National Renewable Energy Laboratory (NREL). (2025, December 10). *The timeline chart of the highest research cell efficiencies for Crystalline Si Cells photovoltaic technologies from 1975 to 2025, (NREL)*. https://www.nlr.gov/images/libraries/pv-images/cellpveffcrysi.webp?Status=Master&sfvrsn=13414312_9

[10] Navarro, E., Costa, N., & Pereira, A. (2020). A systematic review of iot solutions for smart farming. *Sensors (Switzerland)*, *20*(15), 1–29. https://doi.org/10.3390/s20154231

[11] Raj, M., & Prahadeeswaran, M. (2025). Revolutionizing agriculture: a review of smart farming technologies for a sustainable future. In *Discover Applied Sciences* (Vol. 7, Number 9). Springer Nature. https://doi.org/10.1007/s42452-025-07561-6

[12] Rumbayan, M., Abudureyimu, A., & Nagasaka, K. (2012). Mapping of solar energy potential in Indonesia using artificial neural network and geographical information system. *Renewable and Sustainable Energy Reviews*, *16*(3), 1437–1449. https://doi.org/10.1016/j.rser.2011.11.024

[13] Setyawan, A., & Sutandi, T. (2020). Off-grid solar power plant for refrigeration system: A case study in Bandung, Indonesia. *IOP Conference Series: Materials Science and Engineering*, *830*(4). https://doi.org/10.1088/1757-899X/830/4/042031

[14] Sharma, H., Haque, A., & Jaffery, Z. A. (2019). Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring. *Ad Hoc Networks*, *94*. https://doi.org/10.1016/j.adhoc.2019.101966

[15] Silalahi, D. F., Blakers, A., Stocks, M., Lu, B., Cheng, C., & Hayes, L. (2021). Indonesia's vast solar energy potential. *Energies*, *14*(17). https://doi.org/10.3390/en14175424

[16] Sokol, M., Galajda, P., Jurik, P., Pribula, F., & Sokolova, Z. (2024). Design and Implementation of a Wireless Sensor Network Based on the ESP32 for IoT Applications. *Proceedings Elmar - International Symposium Electronics in Marine*, 69–73. https://doi.org/10.1109/ELMAR62909.2024.10694153

[17] Tukiainen, M. (2026, March 24). *Gaisma Sun Path Bandung - crop*. https://www.gaisma.com/en/

[18] Villa-Henriksen, A., Edwards, G. T. C., Pesonen, L. A., Green, O., & Sørensen, C. A. G. (2020). Internet of Things in arable farming: Implementation, applications, challenges and potential. In Biosystems Engineering (Vol. 191, pp. 60–84). Academic Press. https://doi.org/10.1016/j.biosystemseng.2019.12.013

Development of a smart reflex measuring fencing target

Borisz VÁRKONYI[1], Viktor ERDÉLYI[1]
[1]Department of Mechatronics, Institute of Technology,
Hungarian University of Agriculture and Life Sciences, MATE Gödöllő

Abstract

In professional sports, measuring performance—especially speed and reaction time—has become essential. Since speed is a key part of fencing, we built a device that tracks actions with millisecond precision. Because the accuracy of the blade's tip is also vital, the system shows exactly where each hit lands.
The hardware uses a Raspberry Pi computer, and the software is a web application built with the Flask (Python) framework. The device supports four training modes: direct thrust, beat-attack in quarte, beat-attack in sixte, and stay-in-place. Data is sent to the app in real-time, giving fencers and coaches instant feedback. The system is stable and easy to expand, with plans for future upgrades like wireless data transfer and optical sensors.
Field testing demonstrated high system reliability, with 95% of sensor data successfully recorded and negligible signal latency. The microswitch matrix effectively provided precise hit localization, while the integrated statistics module enabled tracking of reaction times and accuracy. User feedback confirmed that the interface is intuitive and the physical construction meets professional durability standards, making the system well-suited for professional training environments.

Keywords

Fencing-specific measurement system, Reaction time and hit localization, Multiplexed microswitch matrix, Embedded systems in sports, Real-time performance analytics

1. Introduction

This chapter discusses the vital role of innovative digital measuring devices in modern sports and highlights the specific need for instant feedback to track a fencer's progress and skill development.
Nowadays, innovative digital measuring devices play a vital role in developing an athlete's skills [1]. These tools allow us to predict more accurately whether a specific training method is effective [4]. In fencing, the development of reflexes, speed, and precision is exceptionally important, leading sports scientists to create diverse training forms to improve these skills.

The main challenge is that we often cannot tell immediately if a training session was successful. Typically, we must wait until the next competition for results [2]. If we only realize a training method was ineffective during a match, it is already too late. While general devices like pulse oximeters exist, they do not always provide relevant data regarding specific athletic performance. This highlights the need for sport-specific measuring devices that provide athletes and coaches with instant feedback on their progress.

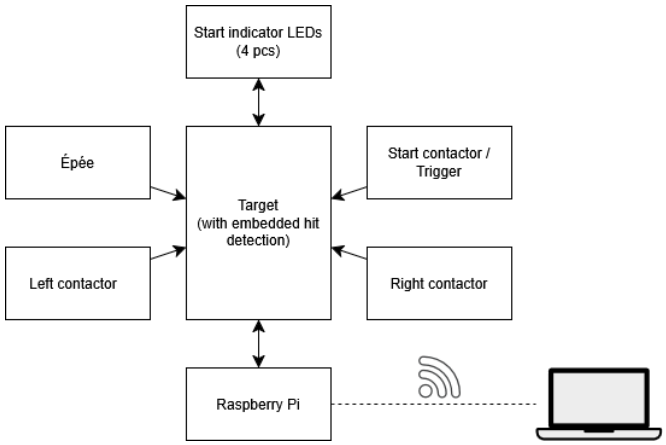
2. Experimental

This section provides a detailed description of the system's hardware components, including the target pad, specialized sensors, the athlete's weapon, and the Raspberry Pi unit used for data acquisition.

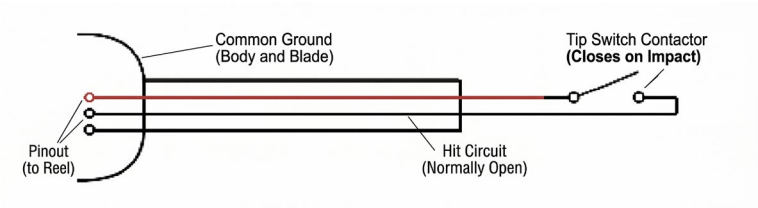
2.1. System component description

The system is constructed of the following hardware components as seen on Figure 1:

- Target Pad: The primary unit for hit detection. It is positioned in front of the athlete and houses the start indicator LEDs. It also contains the embedded sensor subsystem for hit placement detection.
- Left & Right Contactors: Sensors used to detect parries on the Quarte (inside) and Sixte (outside) sides. These consist of two decommissioned épées mounted on either side of the target pad.
- Start Contactor: A sensor (repurposed épée on a tripod) that detects the athlete's starting position. It is the primary trigger for initiating reaction time measurements.
- Épée (Weapon): Figure 2 shows the handheld device used by the athlete. Its electrical design allows for dual-purpose sensing: the blade/body acts as a common ground, while the tip contactor (a normally open switch) closes the circuit upon impact to register hits.
- Raspberry Pi (SBC): The central processing unit responsible for data acquisition and controlling the indicator LEDs. It hosts a web server for user interaction, data visualization, and athlete management [5].



1. Figure - System Architecture Diagram



2. Figure - Electrical Schematic of the Épée

2.2. Physical Construction of the Measuring Device

The device is constructed of the following physical components as shown on Figure 3:

- The device features a multi-layered design built on a custom-cut wooden base, which holds the sensor PCB assembly.
- Protective Layers: A polyfoam layer covers the microswitch-based sensor grid to provide impact protection and dampen hits, preventing damage to the épée tip. This layer also ensures consistent hit registration by transferring pressure to the nearest switch even during off-center impacts.
- Outer Shell: The unit is encased in a treated leather cover identical to the material used in fencing masters' protective gear for high durability. The assembly is secured by a metal frame that holds the leather in place and provides a professional finish.

- External Interfaces: Four top-mounted apertures house the indicator LEDs.
- Two foldable brackets on the front support the épées used for parry detection.
- Cable Management: All electronic components and wiring are located on the rear. Cables for the start sensor and the athlete's weapon are routed through dedicated openings at the bottom of the metal casing.

3. Figure – Physical Architecture

2.3. Multiplexed Microswitch Matrix

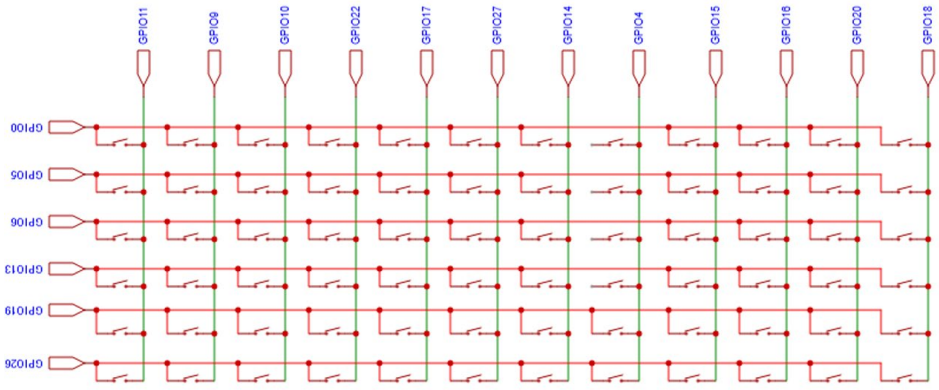
The hit detection system utilizes a 6x12 multiplexed switch matrix located at the center of the target pad, integrating a total of 72 microswitches.

To optimize hardware resources, the system employs multiplexing instead of individual GPIO pin assignments. This reduces the required GPIO count from 72 to 18 pins.

Operating Principle:

- The controller sequentially sets each column to a HIGH logic level.
- Simultaneously, it scans the rows for an incoming HIGH signal.

- When a switch is pressed, the intersection of the active column and the triggered row allows the software to calculate the exact coordinates of the hit.



4. Figure - Multiplexed Microswitch Matrix

2.4. Alternative Hit Detection Experiment: Capacitive Sensing

The initial phase explored a capacitive sensor for hit localization. The design was based on the parallel-plate capacitor model which is described by Equation 1:

$$C = \varepsilon_r \varepsilon_0 \frac{A}{d}$$

Where:

$C - capacitance$

$\varepsilon_r - relative\ permittivity$

$\varepsilon_0 - vacuum\ permittivity$

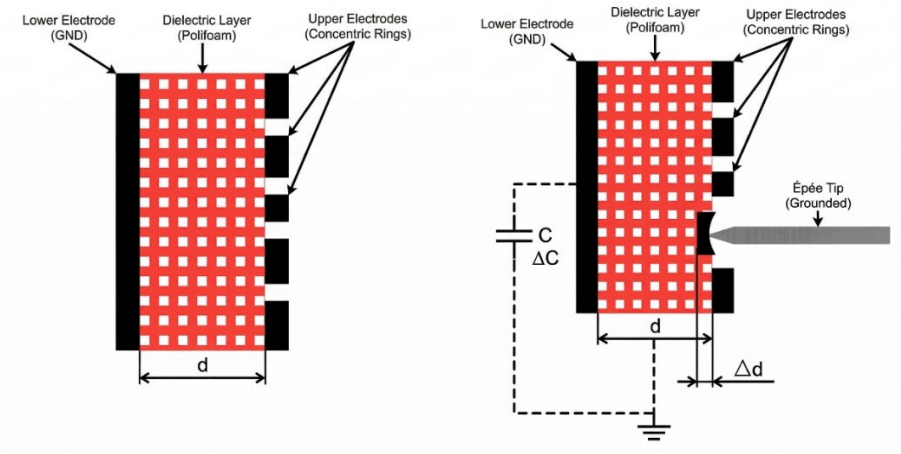
$A - area\ of\ plates$

$d - distance\ between\ plates$

- Construction: A grounded copper base (bottom electrode) was separated from concentric copper rings (top electrodes) by a polyfoam dielectric layer.
- Mechanism: Impact reduces the distance (d) between plates, increasing the capacitance (C).
- Data Acquisition: Due to the lack of analog inputs on the Raspberry Pi, an ESP32 microcontroller was used, utilizing its dedicated capacitive

touch pins and a custom C-based firmware to identify the highest capacitance change.

- Conclusion: The experiment was discontinued because the metallic épée blade acted as a conductor, introducing significant capacitive noise and interference, which led to unreliable hit detection.



5. *Figure* - Capacitive sensor in its idle state & during impact

2.5. Alternative Hit Detection Experiment: Computer Vision

Another approach explored for hit localization was Computer Vision (CV), utilizing the following methodology:

- Reference System: The system used AprilTags (2D barcoded markers) to define the target pad's plane and spatial orientation through specialized detection libraries.
- Tracking: To track the weapon, a high-visibility marker was placed on the épée tip. Blob detection (OpenCV) was then used to identify this marker in the video feed.
- Coordinate Transformation: Using homography (perspective transformation), the detected 2D image coordinates were projected onto the target's physical plane to calculate real-space coordinates.

Conclusion: The method proved impractical for this setup due to two main factors:

- Hardware Limitations: The Raspberry Pi lacked the computational power to perform these operations in real-time.

- Accuracy vs. Cost: Low-cost cameras provided insufficient resolution for precise tracking, while the high-speed, high-resolution cameras and powerful workstations required for a reliable system would significantly increase project costs.

2.6. Software Architecture and Implementation

The software environment is built on a Python-based Flask web framework [7], integrating database management, hardware control, and data visualization.

2.6.1. Data Management (SQLite & SQLAlchemy)

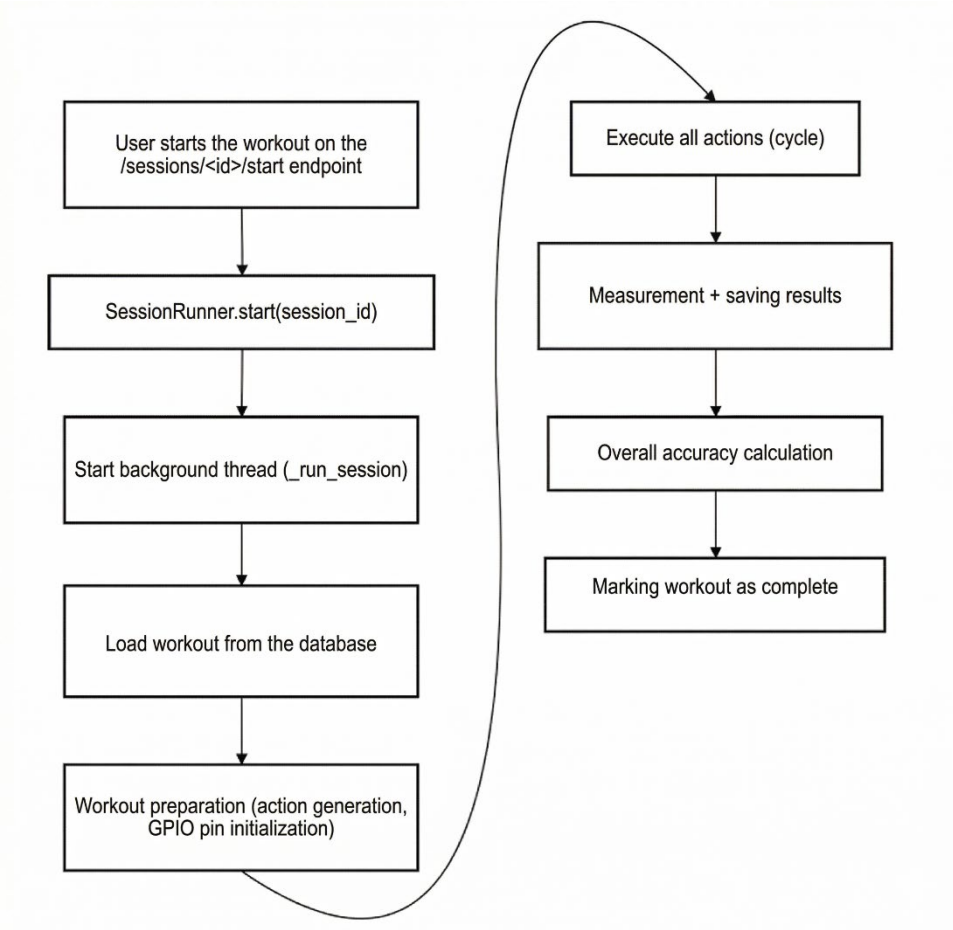
The system uses an SQLite relational database for its low resource footprint and local storage efficiency on the Raspberry Pi [6].

- Schema Design: The database consists of interconnected tables: ATHLETE (profiles), TRAINING_SESSION (metadata), TRAINING_TYPE, and ACTION (predefined exercises).
- Result Tracking: Individual hits and performance metrics (reaction times, coordinates) are stored in JSON format within the SESSION_ACTION_RESULT table, allowing for flexible data structures.
- Integration: SQLAlchemy ORM bridges the database with Flask, enabling dynamic content rendering in HTML templates (e.g., listing athletes or displaying session results).

2.6.2. Core Logic: SessionRunner and HardwareController

The backend logic is divided into two primary functional components:

- SessionRunner: Executed on a background thread, this function manages the training lifecycle. It handles session initialization, randomizes exercise sequences, triggers the hardware, measures timing (reaction, strike, etc.), and calculates accuracy before finalizing the session status.
- HardwareController: Acts as the abstraction layer for the physical device. It manages GPIO initialization, controls the signaling LEDs, and polls the sensor matrix. It ensures a "clean" hardware state by resetting all pins after each session.



6. *Figure* - Flowchart of the SessionRunner function execution

2.6.3. Data Visualization and Statistics

A key feature of the software is the automated statistics generation module. By querying stored session data, the system provides:

- Performance Trends: Graphs showing improvements in reaction and strike times over time.
- Accuracy Metrics: Success rates categorized by action type (e.g., straight thrust vs. parry-riposte).

Hit Mapping: A visual heat map generated from the 6x12 sensor matrix, identifying the athlete's typical hit placement and precision.

3. Results and discussion

This chapter evaluates the practical performance and reliability of the developed system, confirming that the initial research objectives for temporal measurement and hit localization were successfully met.

3.1. System Performance in Practice

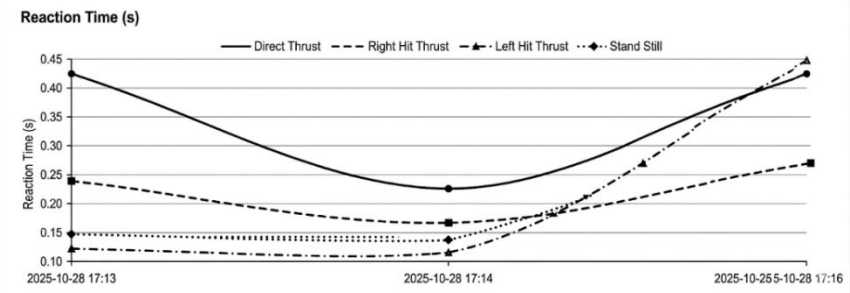
During field testing, the system demonstrated high reliability, with 95% of sensor data being recorded without errors. The visual feedback system performed excellently; the start indicator LEDs provided sufficient brightness and clarity, ensuring that athletes could instantly identify the required task. User feedback regarding the interface was overwhelmingly positive, with participants highlighting the intuitive navigation of the web-based dashboard. Furthermore, the physical construction of the target pad met professional aesthetic and durability standards, comparing favorably to existing commercial sports measurement devices.

3.2. Evaluation of Research Objectives

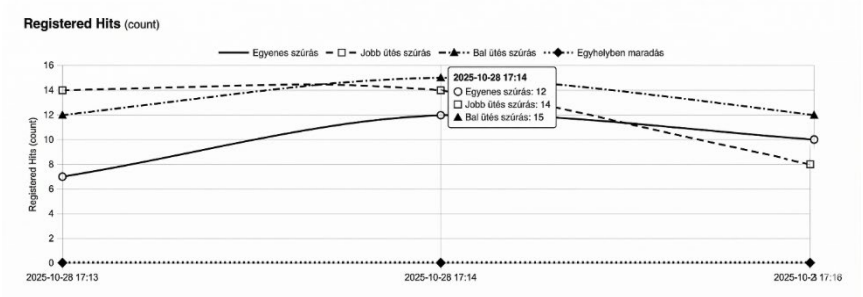
The primary goal of this project was to develop a specialized sports measurement device capable of recording complex fencing actions. The following objectives were successfully met:

- Temporal Measurement: The system accurately measures reaction time from the start signal, as well as the duration of parries (Quarte and Sixte) and the interval between parry and riposte (thrust).
- Hit Localization: While capacitive sensing and computer vision were explored as alternatives, the final microswitch matrix successfully provides precise hit placement data.
- Software Ecosystem: A modern, localized web interface was implemented, allowing for seamless session management and data interaction within a local network.

Progress Tracking: The integrated statistics module fulfills the requirement of providing long-term developmental data and performance analytics for athletes.



7. Figure - Evolution of Reaction Times over Training Sessions



8. Figure - Total Hit Count Registered by the Sensor Matrix

3.3. Future Development and Limitations

Despite the project's success, several areas for technical improvement were identified:

- Sensor Durability: The current microswitch matrix represents a potential point of failure. These switches are not specifically designed for the high-impact forces of repeated épée thrusts. Furthermore, their current integration requires a full disassembly of the target pad for maintenance. Future iterations should explore piezoelectric sensors or industrial-grade force-sensitive resistors (FSRs) for increased longevity.
- Wireless Integration: To enhance the athlete's freedom of movement, the physical cables connecting the weapon and the start contactor should be replaced. The "start" position could be detected via an Inertial Measurement Unit (IMU) or accelerometer mounted on the weapon, eliminating the need for a physical tripod-mounted sensor.

- Connectivity: Moving toward a wireless communication protocol (such as Bluetooth Low Energy or Wi-Fi) between the épée and the Raspberry Pi would significantly modernize the system and reduce the wear and tear associated with traditional fencing body cords.

4. Conclusions

The final chapter summarizes the project’s achievements in creating a functional, web-based measurement tool and outlines the potential for future enhancements like wireless connectivity.

The objective of this thesis was to develop a measurement system capable of accurately recording thrusts and hits in fencing. The primary goal was to create a user-friendly, reliable tool to assist athletes in measuring their reaction times and hitting precision.

The implemented system consists of Raspberry Pi-based hardware and a web-based software application built with the Flask framework. The device handles various fencing actions—such as straight thrusts and parries on the Quarte and Sixte sides—transferring measured data in real-time to a digital interface. The software logs and visualizes this data, enabling coaches and athletes to review results and track developmental progress effectively.

During practical testing, the system demonstrated high stability; the sensors reliably detected hits, and signal latency remained negligible. The user interface is clear and responsive, ensuring a convenient and straightforward user experience. Based on these results, the developed system is well-suited for a professional training environment.

In conclusion, this project has resulted in a fully functional and highly scalable measurement system. Future development could focus on integrating wireless data transmission and expanding the sensor array to further enhance the system's capabilities.

References

[1] [1] Zhao J, Yang Y, Bo L, Qi J, Zhu Y. Research Progress on Applying Intelligent Sensors

[2] in Sports Science. Sensors. 2024; 24(22):7338. https://doi.org/10.3390/s24227338

[3] [2] Géczi, Gábor, Sáfár, Sándor, Havanecz, Krisztián (2023) A kosárlabda-, kézilabda- és

[4] labdarúgás-sportteljesítményt meghatározó változók vizsgálata: Fókuszban gps- és lps-

[5] rendszerek Az Eszterházy Károly Egyetem tudományos közleményei (52. köt., 2022).

[6] Tanulmányok a sporttudomány köréből = Acta Universitatis de Carolo Eszterházy

[7] Nominatae. Sectio Sport. pp. 73-81.

[8] [3] Cadence. (2024, June 14). The history of electronics in sports. Cadence Community –

[9] Corporate News. Retrieved October 22, 2025, from

[10] https://community.cadence.com/cadence_blogs_8/b/corporate-news/posts/the-history-of-

[11] electronics-in-sports

[12] [4] Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G., & Whittlesey, S. N. (2014).

[13] Research methods in biomechanics (2nd ed.). Human Kinetics.

[14] [5] Mathe, S. E., Kondaveeti, H. K., & Vappangi, S. (2024). A comprehensive review on

[15] applications of Raspberry Pi. Computer Science Review, 52(10), 100636.

[16] https://doi.org/10.1016/j.cosrev.2024.100636

[17] [6] Bi, C. (2009). Research and application of SQLite embedded database based on ARM-

[18] Linux. In Proceedings of the 3rd WSEAS International Conference on Circuits, Systems,

[19] Signal and Telecommunications (CISST'09). Zhejiang Wanli University, Ningbo, China.

[20] http://www.zwu.edu.cn

[21] [7] Vyshnavi, V. R., & Malik, A. (2019). Efficient way of web development using Python

[22] and Flask. International Journal of Recent Research Aspects, 6(2), 16–19. ISSN: 2349-7688.

[23] [8] Abdavi, F., Khodadadi, M. R., Heidari Param, A., & Pashaie, S. (2020). Designing the

[24] prototype of smart athletes recording equipment based on Internet of Things using the

[25] Arduino board. Journal of Advanced Sport Technology, 3(2), 42–49.

[26] https://jast.uma.ac.ir/article_882_0bd625accda81b3e3ec78769a905f552.pdf

Solutions for digitizing the human foot orthopedic purposes (brief overview)

László ZSIDAI[1], József NAGY[2], Attila LÁGYMÁNYOSI[2]
[1] *Department of Materials Science and Engineering Processes, Institute of Technology, MATE, Gödöllő,*
[2] *Department of Engineering Informatics, Institute of Technology Hungarian University of Agriculture and Life Sciences MATE, Gödöllő*

Abstract

A significant target group of orthopedic examinations are patients born with central nervous system damage or partially paralyzed as a result of a stroke. The first step in the development of orthopedic aids (e.g. insoles) made for them is sampling, which is typically still traditionally done using plaster sampling. The quality of plaster sampling depends largely on the professional knowledge of the orthopedic specialist and the limb vibrations caused by nervous system damage, which together can hinder the successful manufacture of the device. In the course of their work, the authors review the geometric examinations used during orthopedic examinations in the framework of the application NUMBER 2018-1.3.1-VKE-2018-00013 "Boka-láb ortézisek digitalizált gyártási technológiájának szakértői rendszere/Expert system for the digitalized manufacturing technology of ankle-foot orthoses" to establish its basis. As part of our work, we describe the origin of orthopedic geometric examination tools, then the most important leg-sole deformities, the complex examination of which is the goal of the project. In our description, in a preliminary overview, we propose a recently developed new type of 3D scanning (photometry), the suitability of which is evaluated in comparison with other methods. At the end of our work, we outline goals to be solved to establish the basis for the development of technical tools belonging to the sub-area of the project.

Keywords

orthopedics, foot measurements, 3D scanning, ortho scanners, foot deformities

1. Introduction

The number of people with some kind of musculoskeletal disorder is very high worldwide. Of these, the number of people born at birth with central nervous system damage (cerebral palsy, CP) is significant (~0.2-0.3%). (Steve, 2025) The damage of 1/3 of these children is so great that they need continuous medical and other special treatments, so the burden on society is reduced if the

self-sufficiency, walking ability and working ability can be maintained. The condition of children with cerebral palsy deteriorates as they grow, and as a result of inadequate treatment, many otherwise moveable children live in wheelchairs by adulthood. With proper orthosis care, this trend can be reversed, stopped, or at least the deterioration of the condition can be reduced, depending on the severity of the cases. Correct equipment care significantly reduces the number of orthopedic surgeries required, so it is an important social issue to achieve the best possible condition.

In addition to children with cerebral palsy, there is also a significant number of people suffering from paralysis as a result of stroke. In Hungary, about 40-50 thousand people suffer a stroke every year (aviancare.hu), 48% of whom survive with unilateral paralysis (Karra at. al. 2025), for whom the provision of periodic or continuous aids is a vital necessity for returning to society.

Today, orthopedic appliances for the general patient population in Hungary are made by orthopaedic technicians, where the design of appliances made using the traditional method depends to a large extent on the individual sense, knowledge and experience of the professionals. Continuous limb twitching is a major problem in the case of people with nervous system (even mental) injuries, which greatly complicate sampling.

One of the outstanding goals of the project is to develop a digitization method that replaces the traditional production of assistive devices, which depends to a decisive extent on the individual qualities of the orthopaedic specialist. In the future, instead of plaster sampling, plaster casting and manual carving, as a result of 3D scanning, digital modification and selection of an existing plaster sample/creation of a new plaster sample, production phases will be shortened, the use of auxiliary materials will decrease, and the proportion of incorrectly made tools will decrease, so the results of the project will allow for more efficient and precise work. Fast, clean and painless sampling, faster preparation, more precise tool creation with fewer modifications. In addition to the above, data management (data storage) is also modernized with the use of 3D technology, and even remote sampling becomes possible.

In accordance with the above, our present work prepares for the implementation of the application, reviews and evaluates the opportunities provided by the existing orthopaedic digitization methods.

2. Experimental

The two main aspects that we will follow in our work corresponding to the application subfield are the following:

- The background of the existing and experimental technologies and methodologies required for the research
- analysis and examination of their applicability.

2.1. Basics of foot deformities and the historical background of their study

The examination of the deformity of the foot and sole became intensive with the use of footwear and shoes. Therefore, the instrumental measurement of the foot and sole began with the determination of the fitting of the shoes to the foot and the individual characteristics experienced there (primarily, deviations, deformities). We have been using personalized shoes for centuries, the basis of which is the accurate physical measurement of the foot. This is essential for its comfort design. In the event that the shoe does not match the characteristics of the foot, this can lead to lower musculoskeletal problems such as: foot pain or deformity. There are several types of foot deformities, the following figure shows some of them. (Figure 1.)

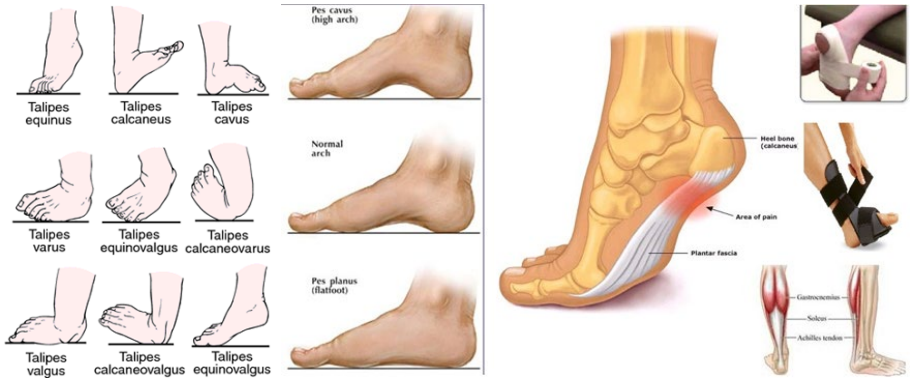


Figure 1. Harmful abnormalities of the sole (Source left: medical-dictionary.thefreedictionary.com), (source right: onlinelearningtips.com)

2.1.1. Foot deformities

In the following, we will briefly review the main changes of the foot and sole (deformation in the next). (Source: dohmeier.hu)

- Flat feet:

Flat feet can be caused primarily by a weak connective tissue structure. Overloading the forefoot is, for example, as a result of inappropriately designed shoes, high heels, overweight, overload due to occupation, long standing. Flat feet can also be a consequential phenomenon by the case of high arches of the feet, ankle prolapse or polyarthritis. (Figure 2, Left side)

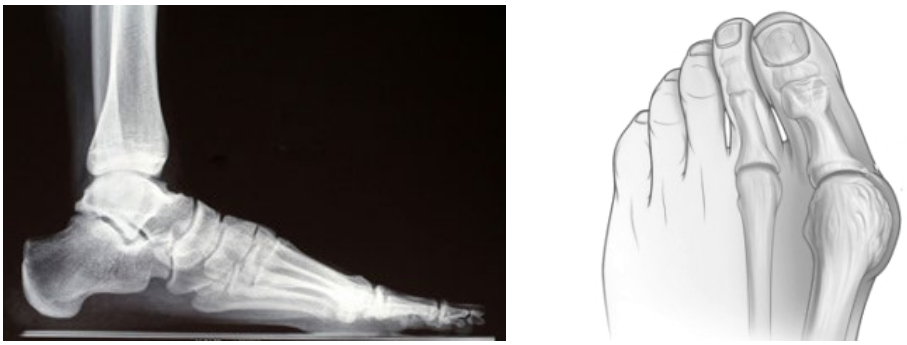


Figure 2. Left: Flat feet on X-ray, right: Sketch of Hallux Valgus (Source: dohmeier.hu)

Shoe inserts with adequate flat foot support in combination with shoe correction (e.g. butterfly sole) usually provide a good solution for relieving pain. In the case of a stiff goose blade, the care is a footbed and relief from the load. Corrective pressure can be painful and completely ineffective.

- Hallux Valgus

Hallux valgus is one of the most common foot deformities. It is often accompanied by transverse prolapse and the big toe typically turns towards the small toes in the basal joint as well. This is often accompanied by inward rotation of the toes. Due to the incorrect position of the joint, the basal joint changes inflamed, which is accompanied by pain. Hallux valgus is caused by an imbalance in the muscles of the foot, during which the old finger turns inwards, towards the little fingers. Since the tension of the muscle fibers out of balance inevitably further increases the incorrect posture, the leg must be physiotherapy. With special orthoses or inserts, the lesion can be stopped. (Figure 2, right)

- High arch of the foot (Pes Excavatus)

The high arch of the foot (high instep, hollowed-out foot) is characterized by the excessive height of the internal and external longitudinal arches in an unloaded and loaded state. (Figure 3, left) The heel bone stands steeply and is slightly inclined (supinated). The angle of action of the front metatarsal will thus be greater. In order for the forefoot to touch the ground, it would have to rotate in the direction of pronation and move in the direction of the plantar. The front and back parts of the foot are facing each other. Due to the increased rotation of the foot, the fingertip of the big toe is placed below the level of the feet. With pronation, i.e. turning the forefoot upwards, the bunion of the big toe becomes deeper and becomes the main load-bearing surface of the forefoot. Lesions of the hammer and claw fingers gradually develop and scars and cornifications form between the metatarsals. Dorsal flexion is limited by longitudinal changes in the talus. The heel bone is in a steep position. Walking becomes cumbersome and inflexible.

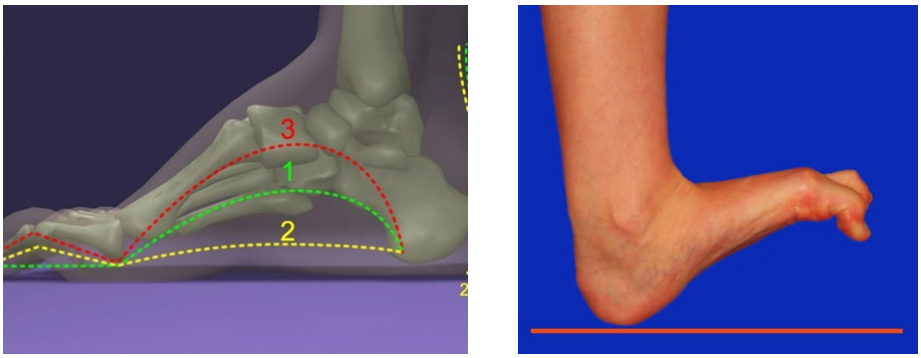


Figure 3. Left: High arch of the foot, 1- Healthy foot, 2- Flat feet (pes planus), 3- Excessively arched, hollowed-out foot (pes cavus) (source: www.easyinsole.com), Right: Pes CalCaneaus) (source: fachgebaerdenlexikon.de)

Steps – inserts to stretch the longitudinal arch. In the case of partially rigid high arches in adults, the provision of insoles is the same as that of the high arch of the foot. However, due to the fixed incorrect position, providing the insoles does not correct the deformity. With a well-sized insert, you can achieve a supportive effect. With this, the pads take over the load at the contact points and relieve the load in accordance with the dynamic holding error. In addition to this, targeted relief of the painful leg areas may also be performed. Additional padding can be used at certain pressure points, such as the cam area. In the case of extreme foot deformities, pain reduction and normalization of the load are often only possible by making an individual footbed. Recesses and cushioning can thus be created much more efficiently.

- "Scythe" Foot (Pes CalCaneaus)

In the case of a scythe foot of the foot, the heel turns sharply towards the ground, thus the function of the upper jumping joint is severely limited. (3rd leg, right) The back of the leg in newborns is almost on the front side of the lower leg. In newborns, a spontaneous restoration of incorrect posture can be observed. Deviation compared to a normal foot: heel bone stands steeply, forefoot cannot be diverted, full load is placed on the heel.

- Hooked foot (metatarsus adductus)

The heel bone is in a suplinated position (tip position) and the arch of the foot is high. The arches are subjected to a strong load. The fingers are clawed, and the leg length is also shorter. (Fig. 4)

Figure 4. Hooked foot "talipes" (source: dohmeier.hu)

It can develop in a hereditary way, with injury and stroke. Here it is important to redistribute the pressure, lateral ankle support, easier rolling down and better gait.

2.1.2. Characteristic curves of the longitudinal arch and heel of the foot

The geometric dimensions of the foot are determined by the measuring points shown in the following figure (Fig. 5).

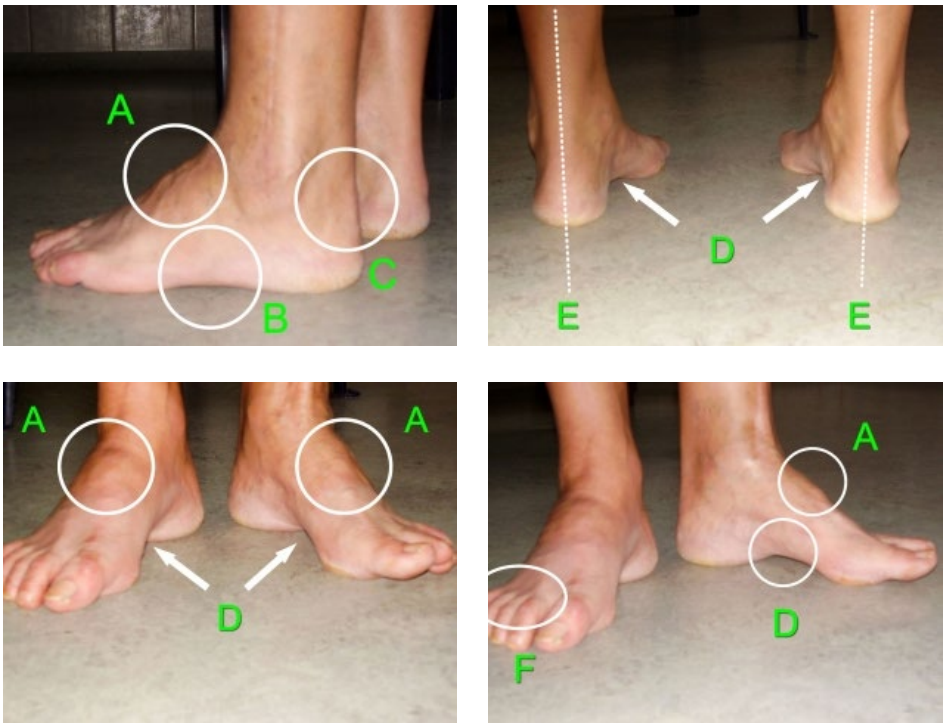


Figure 5. The characteristic curves of the longitudinal arch and heel of the foot. (source https://www.easyinsole.com)

The main geometric features of the "hollowed-out" leg are as follows:
- the heel bone is steep (C), the heel bone is slightly inclined (E) (varus),
- the longitudinal arch of the foot (inner (D) and outer (B)) is morbidly high,
- the back of the foot (A) is also excessively high,
- the toes are in a normal or, more severe case, "sparrow-finger" position (F),
- the big toe often bends towards the inside of the foot (Hallux varus),
- calluses under the toes are common,
- the transverse arch is flattened (F) and the forefoot rests on calluses,
- a smaller or larger bone growth (Haglund exostosis) can be seen in the area of attachment of the Achilles tendon,
- the heel bone is often disproportionately narrow and short,
- under the surface of the heel bone, bone growths (calcaneus exostosis) are often palpable,
- the back of the foreleg is often tilted outwards (supinated),
- the outer side of the forefoot (fingers 4 and 5 and the metatarsus) is located below the level of the forefoot ("twisted under.")

„ An individual with a “hollow” footsteps his shoes strongly outwards, standard half-shoes break the heel of the foot at the bark. Walking in shoes with thin soles is painful. Standard-produced shoes do not want to give the foot the right space, the choice of shoes for individuals is always a strong compromise. These phenomena can be serious.

The determination of congenital and acquired deformities and other foot diseases on the foot and the recommended therapy for application are the task and competence of an orthopedist, rheumatologist, vascular surgeon, neurologist. For therapy using an insole, the selection of the appropriate insole, taking a foot sample, preparing it, and determining the types of shoes that can be recommended for the use of the insole are the tasks of a podiatrist, orthopedic shoemaker, and a specialized workshop licensed to provide services in this field. The design of footwear improves the fit of the shoe based on information regarding the size and shape of the foot. Various measurement methods were used to collect characteristic features of the foot, such as dimensions, foot shape, and sole contour.

2.2. Traditional devices for measuring the feet and soles

2.2.1. Shoe-to-foot fit, vintage tool

The equipment in Figure 6. were also sold as X-rays, pedoscopes or foot-o-scopes from the 1920s in the United States, Canada, England, Germany and Sweden. This wood-covered metal box served the purpose of allowing customers to check the fit of the shoe. The top two holes were for inserting the legs, while the third was for viewing. You could see from both sides. Through this, the location of the foot bones inside the shoe and the outer line of the shoe were clearly visible. The machine was demonstrated by Dr. Jacob Lowe in Boston in 1920 and in 1921 in Milwaukee. (en.wikipedia.org)

Figure 6. The precursor of the geometric examination of the foot is the "shoe/sole" X-ray (source: en.wikipedia.org)

The machine was a great success and thousands of units were sold worldwide. However, this did not last long, because its harmful effects on the human body began to see the light of day. The use of these machines was first regulated in Canada, and later it spread all over the world. The machine could also cause radiation burns, dermatitis, ulceration and carcinoma. (source: en.wikipedia.org)

2.2.2. Common "2D and 3D" device groups for foot-to-foot measurement

Nowadays, even traditional measurements are often used to determine foot deformities, but completely new solutions are also available. The following Figure 7 shows the four main types of currently widespread traditional and modern test instrument groups.

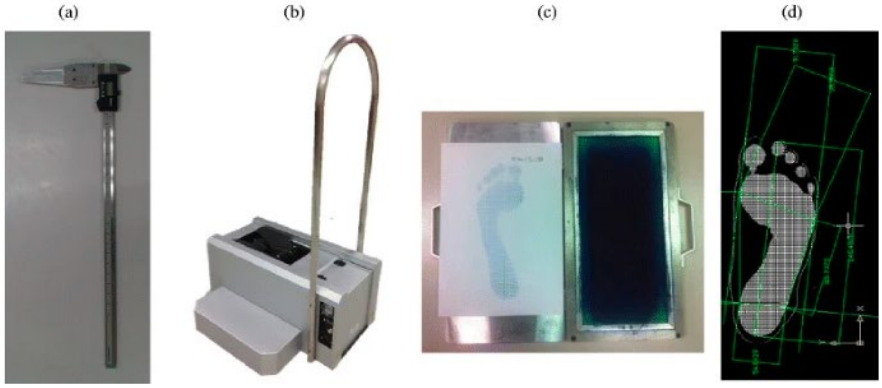


Figure 7. There are four groups of traditional and modern foot measuring devices used today. (Source: https://jfootankleres.biomedcentral.com)

Some of the most common approaches include using digital callipers (vernier) for direct measurements, 3D scanning, and footprint analysis for indirect measurements. Paint layer footprint and digital callipers are the traditional manual approaches used to collect foot dimensions. However, the accuracy of digital calliper measurements is usually affected by human error. Among the 3D scanning methods, the advanced opto-electronic method is gaining more and more ground.

3. 3D Digitization and scanning of the foot

Mapping the shape, size, pressure and movement of the feet. Foot scanning includes everything you need to know about the type of arch, pronation, pressure points, size and even gait analysis. From this you can find out why we feel pain, not only in the feet, but also in other parts of the body.
We can see many 3D foot scans in the literature, essentially using generally developed scanners, mostly for industrial purposes, examples of which can be seen below.

3.1. Application of traditional industrial laser scanners

3D scanners based on laser scanning, which have been used in technical practice for a long time, and the devices developed from them, were the first to appear in various areas of healthcare, including orthopaedics.

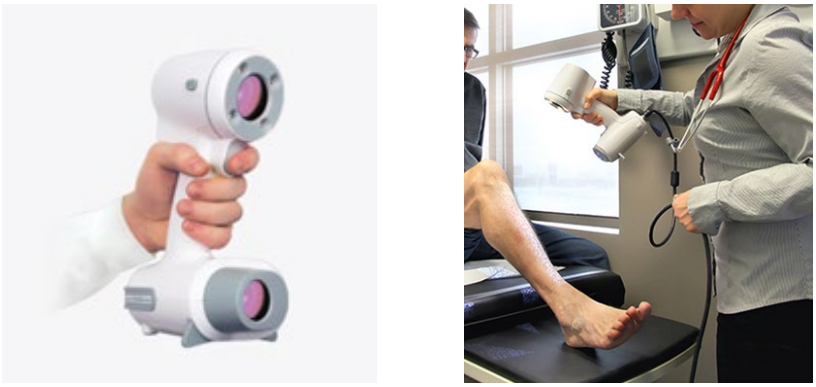


Figure 8. Handheld scanner for healthcare purposes (left image), Scanning a foot with laser scanning (right image), (source: www.creaform3d.com)

The Health Care Partner 3D scanner shown in Figure 8 is an excellent tool for healthcare and 3D body measurement applications. The HCP 3D scanner has

very fast shape recognition (550,000 measurements/second) and high accuracy (0.5mm accuracy). It generates 3D data that is customized and easy to use. The 3D scanner is a great alternative to traditional, tedious and not always reliable measurement methods, such as plaster casts, which is facilitated by the appropriate 96x140x258mm work area size. (2-www.creaform3d.com)

Prostheses for the optical restoration of leg injuries caused by accidents, scanning performed before 3D printing, the resulting model and the model of the prosthesis plan can be seen in the following series of figures 9-11. The common feature of all applications is the use of handheld scanners of different types, but operating on the same principle (laser scanning), and the fact that they are used to prepare prostheses to be manufactured with 3D printing technology.

Figure 9. Scanning for preparation of a leg rehabilitation and the printable model (Source: 3dprintingireland.com)

Figure 10. Scanning a mutilated limb to make a prosthesis (source: 3dprint.com)

The advantage of these devices is precision and a large working space. At the same time, their disadvantage is slowness, which requires a displacement-free model, in our case it is difficult to perform it accurately on a living human leg and requires sufficient cooperation with the patient. This is an obstacle in relation to the project, as children and often mentally or physically disabled patients have to scan their moving, trembling legs. In many cases, color texture is also needed, which these methods cannot solve or can only solve to a limited extent (with subsequent photography). The artec3D product also offers solution related to the above, which can be seen in the following Figure 11. (source: www.artec3d.com)

Figure 11. Digitized model of a child's foot (source: www.artec3d.com)

The 3D shape of the child's leg was created in two minutes with the help of the Space Spider. As you can clearly see, the surface of the leg was properly tracked by the scanner. The anatomical features of the foot, which were in the object, were accurately transferred to the digital model. In terms of direct scanning of the human foot, all organic features would be present in the 3D model, from the heel to the tips of the toes. The former solution is relatively time-consuming (two minutes), so it is worth taking a new direction to implement digitalization, as it requires very fast, "snapshot-like" equipment with medium accuracy and working space. (www.artec3d.com)

3.2. Application of general-purpose mobile smart devices as foot scanners

Health monitoring applications that can be installed on mobile devices are also becoming widespread, among which simple programs/scanners for digitizing the feet have appeared, as can be seen in the following figure 12.

Figure 12. A mobile application suitable for informative foot examinations. (source: jfootankleres.biomedcentral.com)

Foot Management has been in partnership with Sharp Shape for more than two decades. As technology advances, so does foot treatment. The AOMS OCT app was introduced by Sharp Shape, allowing people to scan their feet on iPads. By connecting the sensor shown in Figure 13, iPad can 3D scan the foot. (source: structure.io)

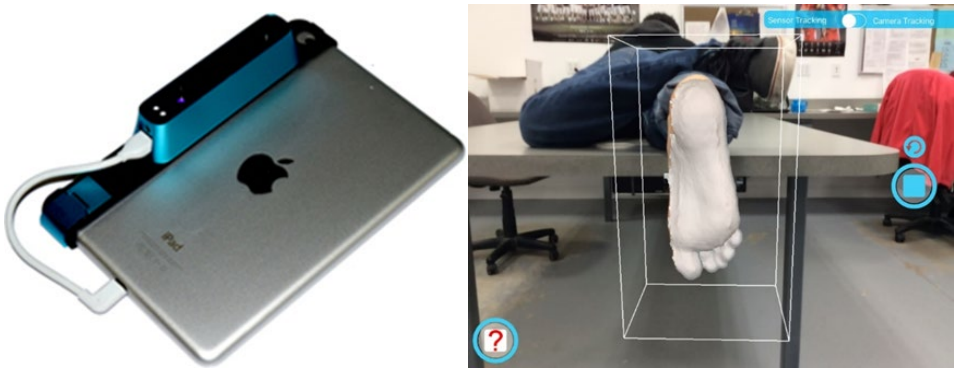


Figure 13. Tablet equipped with an additional scanning head to the left, the same to the right while scanning (source: structure.io)

The possibility that we can use this to make custom orthopedic devices is clear. Traditional scanners are usually bulky and need to be connected to laptops. For true mobility, doctors must be able to perform simple direct mobile scanning, as has already been the case with oral scanners in dentistry. (www.shining3ddental.com)

The system uses E-Z-SCAN ™ for mapping, which is a combination of an Apple iPad, Sensor Scanner and the AOMS OCT app (Figure 14).

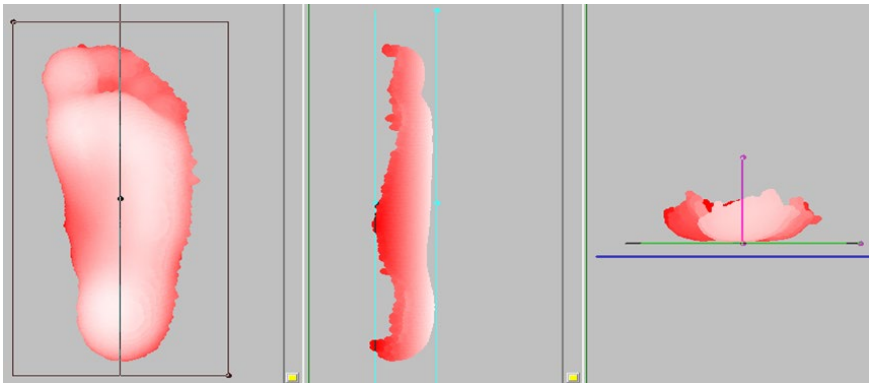


Figure 14. Surface model of the foot scanned with iPad Sense (source: structure.io)

With this system, the doctor can obtain a highly accurate scan of the patient's feet. This can be sent to the distributor, resulting in lower costs for the user. There is no shipping time and it results in greater accuracy than previous scanners. (www.footmanagement.com)

Although these mobile applications are not professional and were developed for home use, for our own orientation examinations, they still show a significant advance, into the use of opto-electronics.

3.3. 3D scanners specially developed for the digitization of feet/soles

In the following, we will review devices that have broken away from traditional industrial applications and have been specially designed for foot scanning. Their workspace, size, scanning accuracy and speed have all been optimized for this purpose.
We can distinguish between devices for sole scanning and devices for whole foot scanning, examples of which can be seen in the following figures.

3.3.1. Foot scanners

Foot scanning records everything you need to know about the arch type, pronation, pressure points, sizing and even provides support for gait analysis. One of the latest foot scanner developments and its operation can be seen in the following figures 15-17.

Figure 15. Cryoscan3D foot scanner, developer: Fuel3D
(source: www.fabbaloo.com)

From the side, the leg is actually slightly pressed into a "suspension membrane". This is a critical element, because otherwise the sole of the foot would be completely flattened, and 3D scans would be of much less value. This is where you can capture the true 3D structure of the foot with Fuel3D's multiple flash technology. (Fig. 16)

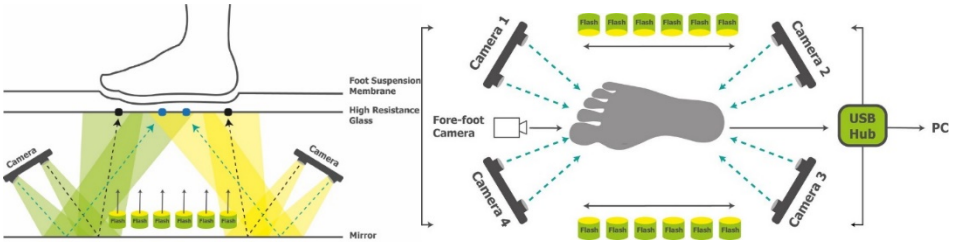


Figure 16. Cryoscan3D process from above view
(source: www.fabbaloo.com)

The Foot Levelers Kiosk (Figure 17) uses graphics and videos to illustrate the relationship between foot imbalances and body balance. This digital foot scanner

highlights imbalances and asymmetries. Foot Levelers orthotics are custom-designed to promote stable pelvic position and balanced posture.

Figure 17. 3D Body View foot scanner (left), (source: www.jassalchiropractic.com), and the associated "Kiosk" (right), (source: www.footlevelers.com)

The Kiosk scanner features that it has a user-friendly touch screen, simple to operate. The posture of the patients and the condition of the soles of the feet can be examined. World-class graphics and videos explain the importance of healthy feet and the benefits of unique orthotics. Exciting new content is added frequently. Cloud-based: Patient outcomes can be securely accessed from anywhere in the world. Comprehensive: You can easily revisit old patient studies to compare them with new ones and record progress. The following Figure 18 shows an example of the displayed images of sole deviations that can be determined on the basis of the test results.

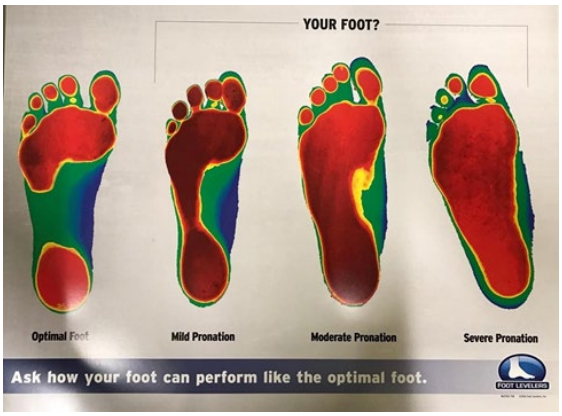


Figure 18. Deformation states displayed during sole scanning (source: www.pinterest.com)

The SC7 is the latest model from Safesize (Figure 19). The 8th generation scanner has been in development for 10 years, in collaboration with HP, Intel and Vitronic. The device can also be used using a smartphone. The app is available on both iOS and Android. The program includes all available dynamic analytics solutions. It is suitable for step analysis with state-of-the-art pressure plate technology.

Figure 19. SC7 Foot Scanner (source: www.safesize.com)

Although the sole scanners presented above show very good results, the results obtained are not sufficient for the project. For us, the "2D" or limited depth 3D scanning of the contact surfaces of the sole itself is not enough. These devices are not suitable for digitizing the complex sole-foot-shin and even the knee. However, they contain useful starting ideas, I am thinking in particular of the development of Fuel3D.

3.3.2. Foot and ankle scanners (Full Foot Scanner)

In the following, we will take a look at specialized 3D scanners that allow scanning the entire foot in addition to the soles of the feet, so-called Full Foot Scanners. In this category, the boxed 3D foot scanners shown in the following figures 20-24 are very common and popular, which scan the foot from several sides with laser scanning inside the box. The reasons for their spread are also the favourable prices.

Figure 20. "In box" full-foot scanners and their application (source: orthotech.com.au) and (source: www.envisicveriscan.com)

The VeriScan Full Foot Scanner used to create the unique insoles in the upper right corner of Figure 20 encompasses the entire foot. In addition to the medial, lateral and plantar aspects, the upper parts of the foot are also digitized. It features a four-camera, true 3D system that scans the entire foot and then outputs a full-point cloud for processing. These systems use multiple cameras and lasers for the lower and lateral parts of the foot, using mirrors to digitize the sole. The following figure 21 shows the Easy-foot-scan box full-foot scanning equipment.

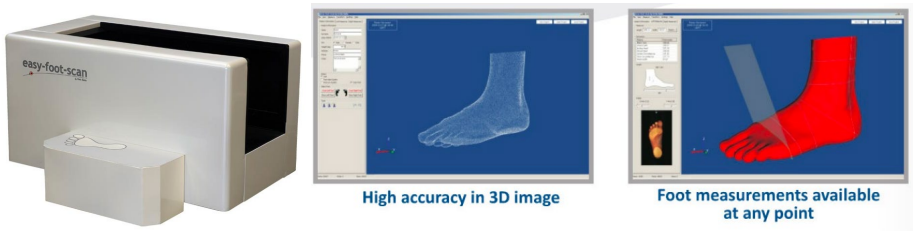


Figure 21. Easy-Foot-Scan equipment and the foot models made with it (source: www.orthobaltic.eu)

Key specifications of the equipment, working area 400x200x 200mm, accuracy (max): ± 0.3mm, scanning speed 5 sec (3D only), 10 sec (3D + 2D).
The Albert scanner shown in Figure 22 collects unparalleled data and information about the legs as it has 18 digital cameras, 960 infrared LEDs and receptors, and 5184 gold-plated sensors. Albert offers one of the most sophisticated, interactive, in-store foot scanning experiences.

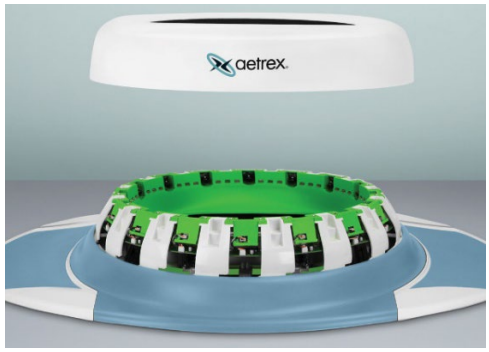


Figure 22. Albert scanner (source: vimeo.com)

The main features of the Albert scanner are the pressure measuring area of 36cm x 36cm, 16 cameras for 3D viewing and maximum data collection, 2 cameras for

pronation analysis, colour LED display. The machine has a pressure sensor plate, with the help of this device, the machine is able to collect data on the pressure points of the sole. The materials collected in this way make it easier for the user to determine the pressure areas of the foot and their size. The machine collects the main characteristics of the foot, such as: length, width, foot diameter, arch height with an accuracy of up to a tenth of a millimeter. From the data collected in this way, it creates a 3D model. On this, the pressure values are indicated. The "Fit HQ" part of the program ensures that the shoe is perfect from the first try. It uses artificial intelligence, which is connected to the retailer's POS or ECommerce system. Through this, the system is able to recommend the best fit. (source: www.aetrex.com)

An interesting solution is True Fit (Tactile) shown in the figures below, which can also scan the entire foot up to the mid-calf. (Figure 23)

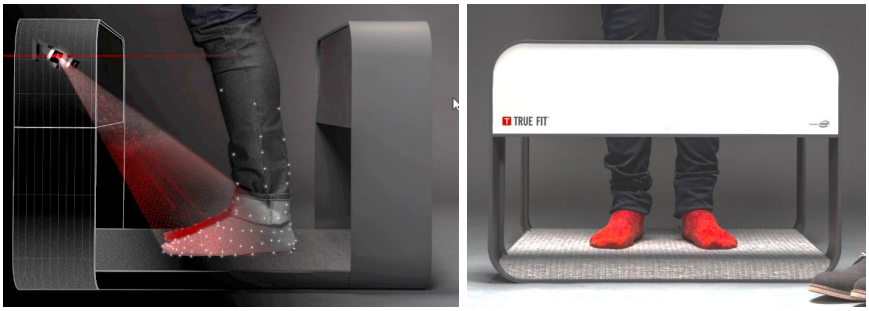


Figure 23. Tactile True Fit foot Scanner (source: tactileinc.com)

Tactile has designed a great prototype with the right strength and design, easy to transport with low weight. Tactile used engineering and design expertise to use CAD simulation to accurately position the cameras. Colour, materials and finishes (CMF) have been developed with durability in mind.

3.3.3. Photometric 3D scanning

Laser scanning, as a scanning method, has a large working area, making it excellent for digitizing stable machines and objects. At the same time, laser scanning is slow, making it less suitable for scanning moving living beings where the operation time is crucial. Scanners operating on the principle of photometry (similar to optoelectronics) work quickly with snapshots, which allows for very fast scanning within seconds.

Photo scanning can also be done from shots taken from different angles of the same object with a digital camera, of course, the advantage of speed is lost here. The previously mentioned Fuel3D company developed (FUEL3D Scanify), which was among the first to offer a suitable solution for the former. With this

equipment, 3D scans of great depth can be made with a single shot, with sufficient coverage, as can be seen in the following figure 24.

Figure 24. Fuel3D Scanify scanner (source: www.dream3d.co.uk)

With the hand-held unit shown in Figure 24, we essentially take a snapshot, and the surface/area that the device can "see" is digitized. Of course, it cannot detect undercut surfaces, so the resulting surface model needs post-corrections (but this does not mean, for example, additional stress on the patient's leg in our case). To spatially align the object and the scanner, a calibration sheet is required (a small black and white circle in front of the object), which must always be placed in the immediate vicinity of the object.

This 3D scanning does not work by scanning the laser line, cross or projected contrast line, zebra stripe and cubes projected onto the surface, but determines the position of the surface points much faster, using the triangulation method performed on the image of the object by two cameras at an angle to each other. This method is similar to "photogrammetry", which can also be considered a newer technique.

The head scanner shown in Figure 25 is interesting, where these handheld units (author's note: the FUEL3D scanify was primarily developed for face scanning, e.g. for designing glasses) are placed on a frame and operate the head scanning in an instant, operating simultaneously with appropriate software coordination. A similar solution may arise in our project.

Figure 25. Head scanning using Scanify scanners in a ring, using them simultaneously (source: 3dprint.com)

3.4. Modern technology for the production of special shoes

After the introduction of the foot scanners, I consider it necessary to look at the modern method of producing orthopedic shoes after foot examinations. Essentially, this is what we lay the foundation for with scanning.

There are two main groups of special shoes: individual shoes and orthopedic shoes. The main differences between the two are:

- Measure or scan the client's feet
- They personalize the shape and style or adapt the characteristics to the client's condition (e.g. in diabetics).
- They create a new style or modify the existing one to suit the individual.
- They print, draw or cut a pattern for manual assembly.

However, the most important difference is that custom shoes are made for clients who do not have more serious deformities. In contrast, orthopedic shoes are developed specifically for people with deformities or diabetics. And the form, design and raw materials are also subordinated to this.

The Shoemaster® Custom program was developed specifically for creating custom shoes (Figure 26). The system provides tools to analyze the shape and size of the legs, and if necessary, you can quickly and easily adjust the most important final measurements. The measurement system allows you to compare your legs by section, and displays the results in 2D or 3D format (Figure 27).

There are many standard sections and measures included in the system with the ability to add user-defined parts if needed. The powerful final adjustment system with the help of the modeling tool kit allows you to quickly and easily change the final shape and size. Settings can be changed in order to make changes in height, latitude, or specific localized areas. (source: www.technorg.hu)

Figure 26. Shoemaster custom technological process for the production of orthopedic shoes (source: www.technorg.hu)

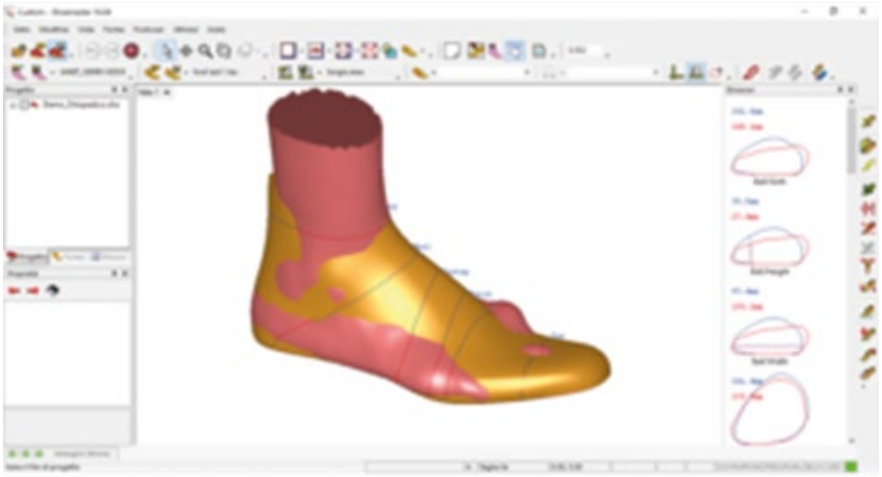


Figure 27. Software for orthopedic modification of the scanned foot model (source: www.technorg.hu)

Conclusions

From the brief presentation of the discussed foot deformities, it became clear that the orthopedic application defined as the project goal places the following main demands on us:

- complex spatial digitization (Full Foot Scanner),
- minimization of external pressure on the surfaces to be tested (e.g. foot rest),
- minimizing scanning time due to mentally handicapped patients,
- easy application,
- mountability and transportability,
- creating a coherent surface model with sufficient accuracy (the dimensional accuracy of mechanical applications is not necessarily the goal).

The presented 3D foot scanning solutions show many differences. In practice, they are used with more or less success, of course, this depends on the purpose of use. In the case of cooperative and patient patients, a hand-held scanning laser scanner adopted from the industry also gives adequate results. The disadvantage is that it is time-consuming and often complicated to create a model, but at the same time it is very accurate and with sufficient support of the foot, there is no need to deform the foot, especially the sole, with external supports. Sole scanners are excellent for mapping the support surfaces of the sole part and smaller arches, they are easy to handle even by the patient, they are also available with fast (snapshot) scanning time (e.g. Cryoscan) and provide fast diagnostics. However, they do not digitize the entire foot or foot, so they do not create a model suitable for complex orthopedics. Due to the weight on the sole, indentations are created on the sole that interfere with the model, so it provides additional information for step analysis and posture assessment.

The presented full-foot scanners (FFS), which are widespread in healthcare, create a continuous surface model of the entire foot. Their accuracy varies depending on the manufacturer, the resolution of older designs is poor, so details are lost from the surface. Scanning is done with laser heads moving along a moving mechanism, which is time-consuming (more than 0.5 minutes). The fairly closed box and the insertion of the foot require a cooperative patient, the sole loaded with body weight is pressed against the lower glass plate of the scanner, thus changing the shape of the sole.

The application of photometry or opto-electronics in 3D scanners is fast (with snapshots) and, unlike laser scanners, is also capable of producing coloured textured images, which increases the diagnostic possibilities of the examination. We have already presented a foot scanner operating with photometry in the lesson, where we also provided several characteristics. However, we are less familiar with a photometric solution that can scan the sole without load, which also scans the entire foot at the same time without causing any disturbance to

mentally impaired patients.

In summary, we see a realistic development opportunity for foot digitization using the photometric 3D scanning principle and appropriate support of the exposed foot, taking into account the project goals. The assembly construction principle, concept presented by TruFit for the construction of the structure to be developed shows a good direction for the construction of the structure's frame.

In accordance with the objectives of the project, two digitization procedures were proposed, which are the following:

- Laser scanning on a plaster sample, which corresponds to industry practice.
- Examination of the possibility of real-time rapid scanning of the patient.

In connection with technical development, the following questions arise:

- Determining the characteristics of the imaging cameras (taking single photos or videos is preferable) and their number (expected to be at least 5).
- Determining an exact trajectory for the placement positions of the imaging cameras. To do this, the cameras must be placed along a trajectory according to a given function that matches the geometric characteristics of the human foot. For the foot dimensions, it will be necessary to prepare several plaster models, the appropriate data can be obtained by 3D scanning them.
- Examination of the overlapping areas of the digitizing cameras and the interfering effects.
- The identification of the photos or video recordings by camera must also be clearly resolved.
- It is important to take short, even instantaneous, recordings to eliminate the disturbing effects of the previously mentioned nervous system problems or limb tremors.

During the implementation, we have to solve the following important tasks:

- Minimizing the number of cameras.
- Calculated precise placement of cameras along the track.
- Implementation of short/instantaneous recordings (e.g. using photometry).

1. Parallel processing of recordings from multiple cameras (5-10), analyzing the data rate.

Acknowledgements

This research was supported by the 2018-1.3.1-VKE-2018-00013 SZÁMÚ „Boka-láb ortézisek digitalizált gyártási technológiájának szakértői rendszere” project (National Research, Development and Innovation Office, Hungary).

References

[1].Aviarance.hu, (https://aviancare.hu/blog/stroke-statisztikak-magyarorszagon)

[2].Karra Geetha, G. Sasanka, Meraj Unnisa Banu, S. Pridvineel Reddy and T. Rama Rao (2025): Cerebrovascular accident associated hemiparesis and hemiplegia: coping strategies and support, International Journal of Pharmaceutical Sciences and Research, IJPSR, 2025; Vol. 16(5): 1329-1334 (https://ijpsr.com/bft-article/cerebrovascular-accident-associated-hemiparesis-and-hemiplegia-coping-strategies-and-support/)

[3].Steve (Open Class Action) (2025): What percentage of children are born with cerebral palsy, Activities, Alzheimer's, HelpDementia.com, September 26, 2025, (https://helpdementia.com/what-percentage-of-children-are-born-with-cerebral-palsy)

[4].http://fachgebaerdenlexikon.de/index.php?id=2140

[5].http://www.envisicveriscan.com/full-foot-scanner/

[6].https://3dprint.com/113991/fuel-3d-ces-2016/

[7].https://3dprint.com/41606/3d-printed-prosthetic-leg/

[8].https://3dprintingireland.com/portfolio/pharmaceutical-medical/attachment/scanmain/)

[9] https://dohmeier.hu/talpbetet/labfejtorzulasok

[10] https://en.wikipedia.org/wiki/Shoe-fitting_fluoroscope

[11] https://jfootankleres.biomedcentral.com/articles/10.1186/s13047-014-0044-7?fbclid=IwAR0w9iIDAPk4jEu12ch5z_Jd47fYzsdmdnF364HVt_VS-ysPeVSr-qW0KOk

[12] https://jfootankleres.biomedcentral.com/articles/10.1186/s13047-014-0044-7?fbclid=IwAR0w9iIDAPk4jEu12ch5z_Jd47fYzsdmdnF364HVt_VS-ysPeVSr-qW0KOk

[13] https://medical-dictionary.thefreedictionary.com/talipes

[14] https://onlinelearningtips.com/2013/10/is-plantar-fasciitis-ailing-you

[15] https://orthotech.com.au/3d-scanners/

[16] https://structure.io/

[17] https://tactileinc.com/work/intel-realsense-foot-scanner/

[18] https://vimeo.com/366345569

[19] https://www.aetrex.com/albert-foot-scanner.html

[20] https://www.artec3d.com/3d-models

[21] https://www.dream3d.co.uk/product/fuel3d-scanify-ex-demonstrator/

[22]https://www.easyinsole.com/hu/extrem_labbetegsegek/magas-labboltozat/34/

[23] https://www.fabbaloo.com/blog/2016/6/7/3d-foot-scanning-now-reality-3d-printed-orthotics-to-follow

[24] https://www.footlevelers.com/innovations/3d-bodyview

[25] https://www.footmanagement.com/foot-scanner

[26]https://www.jassalchiropractic.com/2019/04/introducing-our-digital-foot-

scanner/
[27]https://www.orthobaltic.eu/catalogues/Scanerio_En/mobile/index.html#p=1
[28] https://www.pinterest.com/pin/387098530457266117/
[29] https://www.safesize.com/foot-scanning/#nextsection
[30] https://www.technorg.hu/cad-cam/footwear-cad/custom-made/
[31] www.creaform3d.com/en/health-care-solutions/health-care-partner-3d-scanner
[32]www.creaform3d.com/sites/default/files/assets/brochures/files/hcps_brochure_en_20032012.pdf
[33] www.orthobaltic.eu/footwear/easy-foot-scan.html
[34] www.shining3ddental.com/solution/aoralscan-3/

Motion analysis of a 3D ankle model

László KÁTAI[1], Péter GÁRDONYI[1], István NAGY[1]
[1] *Department of Machine Construction, Institute of Technology*
Hungarian University of Agriculture and Life Sciences MATE, Gödöllő

Abstract

This study presents the development and motion analysis of a 3D biomechanical ankle model. A physical ankle model was digitized using scanning techniques to create a refined solid geometry. The model was processed and prepared for kinematic simulation using appropriate geometric features. Motion analysis was conducted in a multibody simulation environment using defined kinematic pairs. Primary ankle movements, including dorsiflexion and plantar flexion, were investigated. Ligaments were modelled as spring elements based on literature-derived stiffness values. Simulation results provided insight into ligament forces and displacement during motion. The approach demonstrates an effective method for simplified yet realistic ankle joint analysis.

Keywords

ankle joint, Adams view, 3D modelling, motion analysis, kinematical simulation

1. Introduction

Understanding the biomechanics of the ankle joint is essential for improving injury prevention, rehabilitation, and performance in activities involving human locomotion. Due to its complex anatomical structure and the interaction of bones, ligaments, muscles, and tendons, the ankle joint exhibits highly coordinated motion that is difficult to represent accurately using simplified mechanical models. Consequently, the development of reliable biomechanical models has become an important focus in musculoskeletal research.

One of the main approaches used in the analysis of ankle mechanics is multibody modelling. Multibody models enable the representation of the ankle complex as a system of interconnected rigid bodies, allowing researchers to investigate joint behaviour during dynamic activities such as walking, running, or jumping (Marta, G et.al. 2022). Although several models have been proposed, detailed representations of the ankle joint remain relatively limited, and many existing models focus primarily on joint angles and ground reaction forces. Simplified mechanisms, such as single-degree-of-freedom spherical representations, have been introduced to approximate passive ankle motion while maintaining computational efficiency (Sancisi, N. et.al. 2014).

In addition to structural modelling, kinematic and dynamic analyses play a crucial role in understanding ankle joint function. Kinematic data, including joint angles and ranges of motion, provide insight into how the ankle behaves during different phases of movement. Dynamic and kinetic analyses complement these measurements by examining the forces and moments acting within the joint. The integration of kinematic and kinetic information is therefore fundamental for accurately describing the mechanical behaviour of the ankle complex.

Advances in musculoskeletal modelling have further enhanced the ability to estimate mechanical characteristics of the ankle joint and simulate its response under various loading conditions. These models contribute to a deeper understanding of the mechanical demands placed on the joint during everyday and athletic activities. As a result, they support the development of improved clinical interventions, including orthotic design, rehabilitation protocols, and injury prevention strategies (Jamwal, PK. et.al. 2017).

Overall, the modelling of ankle joint motion requires the integration of multibody biomechanics, kinematic measurements, and kinetic analysis. By combining these approaches, researchers can obtain a more comprehensive representation of ankle joint mechanics, ultimately contributing to more effective treatment methods and improved functional outcomes.

By analyzing ankle movements, the following conclusions can be drawn. There are three fundamental planes in the anatomical structure of the body according to Fig.1.. The plane dividing the body into two halves is called the sagittal plane. The plane perpendicular to this plane but running along the body's longitudinal axis is called the coronal plane. The third plane is perpendicular to the previous two planes; this is the transverse plane.

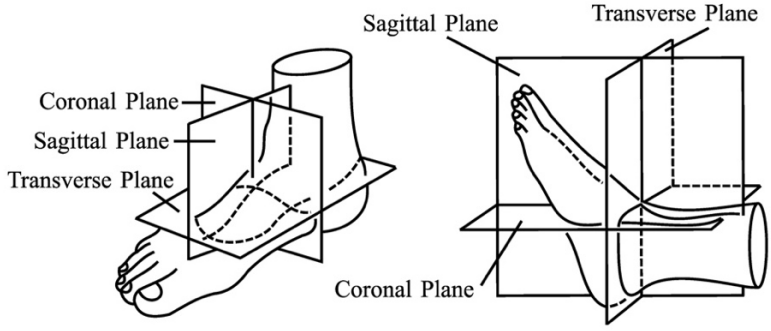


Figure 1. The main planes of the ankle (Hongyu Zhao et.al. 2019)

Ankle movement involves all three planes, so we can interpret different types and directions of movement along these planes. The primary movements are

those we perform most frequently with the foot; in this case, these are raising and tilting the foot, movements known as dorsiflexion and plantar flexion. This movement can be described as a hinge-like motion within the ankle joint. Its axis is parallel to the intersection of the transverse and coronal planes. We can refer to eversion and inversion as secondary movements, which occur around the main axis of the ankle; their axis is parallel to the intersection of the coronal and sagittal planes. We can also perform this type of movement while keeping the foot pressed against the ground, moving the foot outward and inward. In inversion, the outward movement of the foot shifts the sagittal plane outward; in eversion, the sagittal plane shifts inward. The tertiary movements, which occur around an axis defined by the intersection of the transverse and sagittal planes, are called abduction and adduction. Such movements occur in the case of ankle sprains; we cannot maintain our stability during these displacements, as our foot is positioned at a certain angle to the ground [4]. The possibilities for foot displacement described above are illustrated in Figure 2.

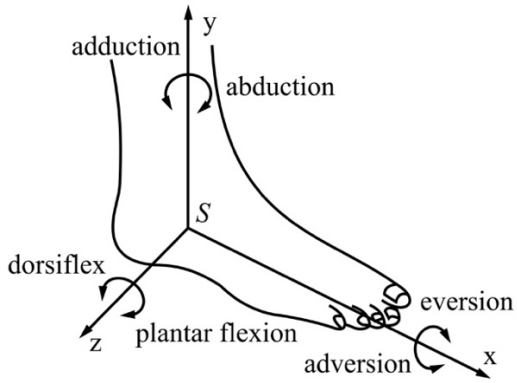


Figure 2. The main axes of motion of the ankle (Ting-Cheng Chang, Xiao-Dong Zhang 2019)

Based on data from the literature, we determined the standard values for dorsiflexion and plantar flexion.

Rotation direction	**Mode**	**Range**
Around z-axis	Plantar flexion	40°-45°
	Dorsiflexion	25°-30°
Around y-axis	Abduction	15°
	Adduction	35°-40°
Around x-axis	Eversion	25°-30°
	Adversion	25°-30°

Figure 3. Values of the maximum displacements caused by the anatomy of the ankle (Ting-Cheng Chang, Xiao-Dong Zhang 2019)

In this article, we will demonstrate how to generate a 3D solid model from a physical model of the ankle, as well as a simplified method for conducting motion simulation analyses.

2. Ankle 3d geometric model development

2.1. 3D solid model creation

To perform the kinematic analysis of the ankle, based on a physical model a 3D anatomical solid-body model were created.

As a first step, the available physical model was scanned. The figure below shows an intermediate stage in the creation of the resulting surface model.

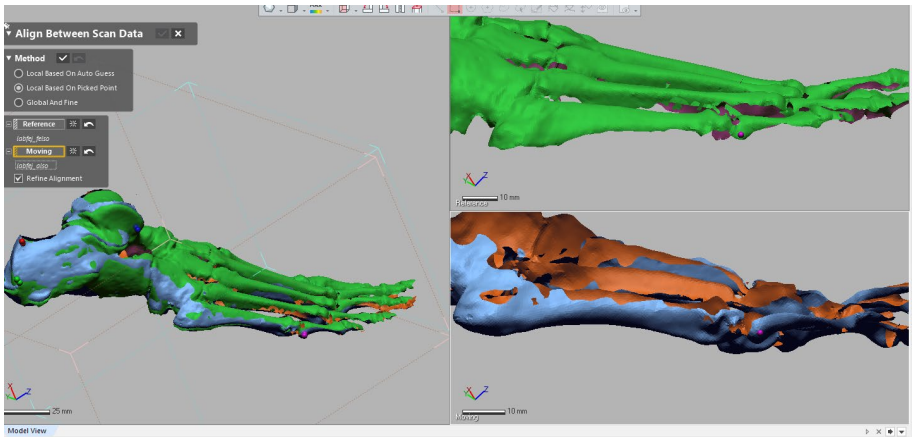


Figure 4. Merging scanned surface model parts

After scanning the physical model, the resulting surface model had to be refined, which involved filling in any gaps in surface continuity and removing parts of the model that were “irrelevant” for further analysis. It was then possible to create a solid model from the surface model delimited by closed surfaces.

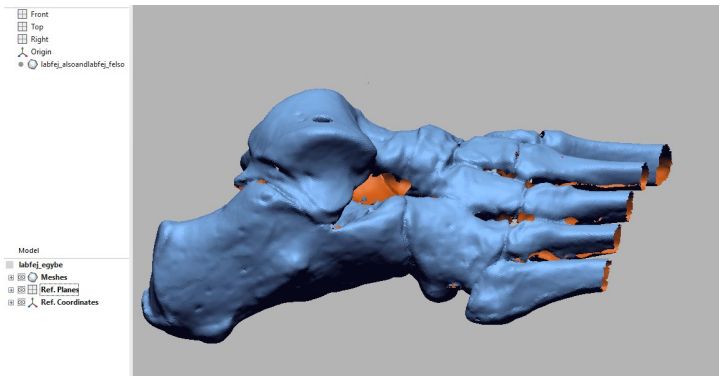


Figure 5. Improving the surface model

The foot model and the lower leg model were created separately.

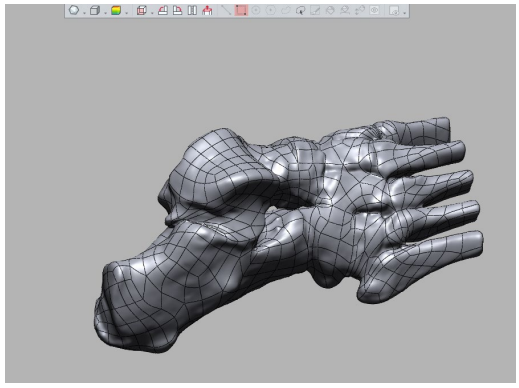


Figure 6. Solid modelling of the scanned foot

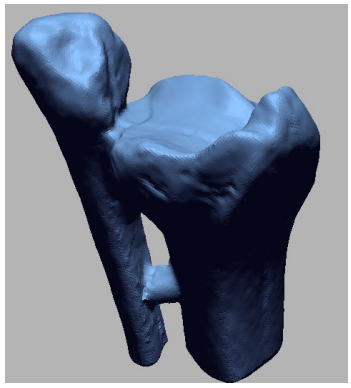


Figure 7. Solid modelling of the scanned leg

In order to properly define the relationships (kinematic pairs) between the two components (lower leg and foot) in the kinematic analysis software, specific geometric features were created in the 3D modeling software. These geometric features are shown in the figure below. The cylinder formed at the end of the lower leg indicates the "vertical" direction, meaning that the lower leg is in its default position relative to the principal planes.

Figure 8. The cylindrical feature indicating the orientation of the initial position

Another design feature indicating the direction of rotation is located on the lower leg, corresponding to the position of the axis of rotation, and thus indicates the imaginary centre of rotation.

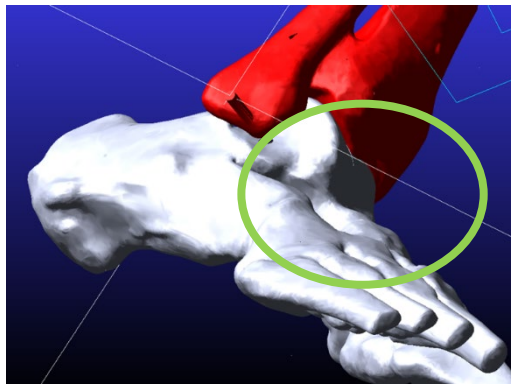


Figure 9. The distinctive design feature on the lower leg

The motion of the ankle model was examined in the sagittal plane. As indicated in Figure 2, this refers to rotation about the z-axis.

2.2. Kinematical model creation in Adams/view

The motion simulation of the ankle model can be performed by exporting the 3D solid model to a suitable file format. Possible formats include IGES, STEP, and Parasolid. After conducting several tests, we determined that our simulation model had the necessary shape characteristics when using the Parasolid file format.

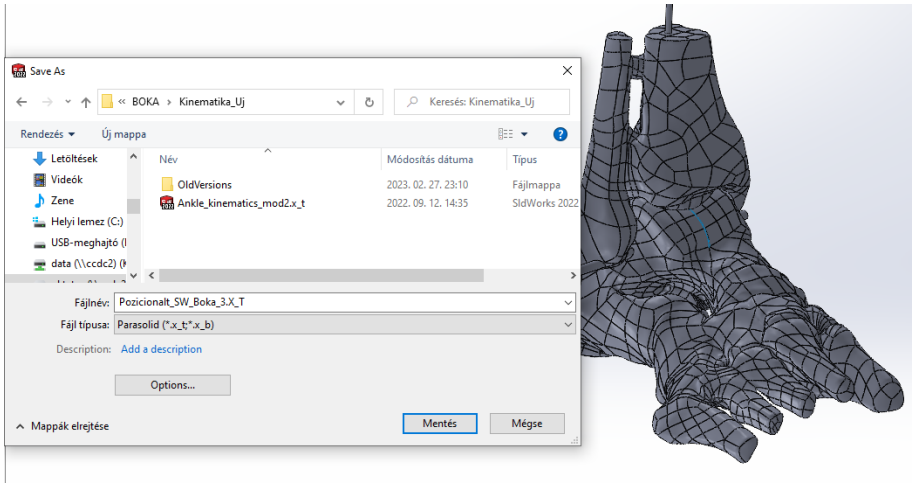


Figure 10. Saving the ankle model in Parasolid format

To implement motion simulation, it is necessary to define the appropriate kinematic pairs. These are as follows:

- Leg – base (ground) – fixed joint
- Foot – lower leg – revolut joint

Adjusting the hinge joint is a challenging task due to the unique shape of the ankle model, as there are no clearly defined, distinguished directions on the solid model created from the scan. As mentioned earlier, we attempted to solve this problem by creating a cylindrical rod element at the imaginary center of rotation (a “geometric feature” indicating this was present on the physical model) during the construction of the 3D solid.

To set the direction of the axis of rotation, we selected the software option that allows us to define the desired axis of rotation perpendicular to the set direction of the grid.

The hinge connection between the two ankle sections was then adjusted.

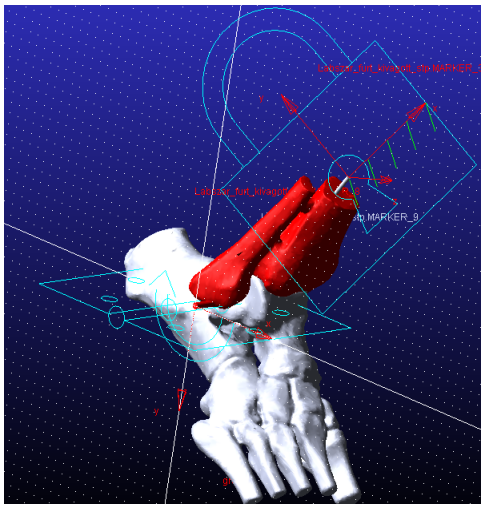
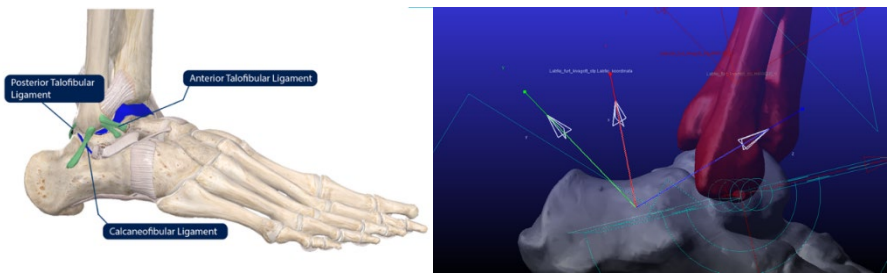


Figure 11. Pairs of kinematic constraints defined in the model

3. Results and discussion

3.1. Setting the motion parameters

We verified the model's operation and set the dorsiflexion angle to 30 degrees. We also verified the rotation at 40 degrees of plantar flexion.
Our additional goal with the motion simulation was to examine the displacement of joint contact points and to determine the forces generated in the joints by adjusting the joint parameters. To determine the displacements, we attached coordinate markers at the joint contact points. We determined their positions based on analyses of the literature.



a) b)

Figure 12. Ankle ligaments (a) (SW KONG 2020),
Defining attachment coordinates for ligaments (b)

We modelled the ligaments by defining them as spring elements. To do this, sources from the literature were used (Fig. 14.)

The spring adjustment parameters for the ligaments are illustrated in the Figure below.

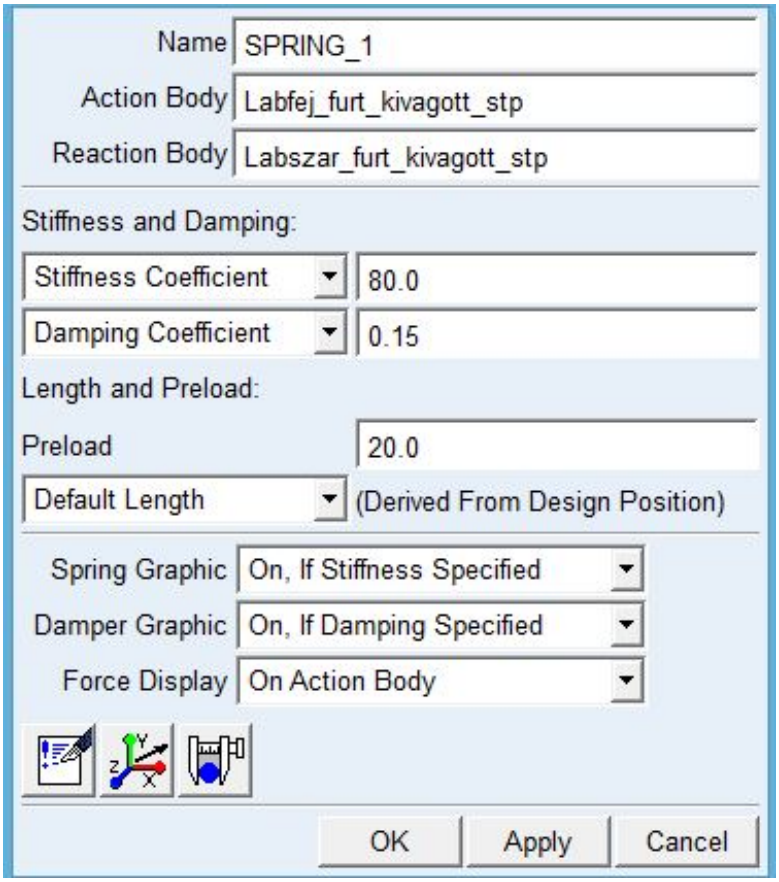


Figure 13. Setting spring parameters

Ligaments represented in the models	Stiffness (N/mm)
Interosseous membrane (4 ligaments)	400
Anterior tibiofibular (distal)	78
Posterior tibiofibular (distal)	101
Anterior talofibular	90
Posterior talofibular	70
Calcaneofibular	70
Anterior tibiotalar	70
Posterior tibiotalar	80
Tibiocalcaneal	122
Tibionavicular	40
Interosseous talocalcaneal	70
Lateral talocalcaneal	70
Medial talocalcaneal	70
Posterior talocalcaneal	70
Dorsal talonavicular (2 ligaments)	70
Calcaneonavicular (dorsal & planar)	70
Calcaeocuboid (dorsal & short planar)	70
Cuboideonavicular (dorsal & planar)	70
Cuneonavicular (dorsal & planar)	70
Intercuneiform (dorsal & planar)	70
Tarsometatarsal (dorsal & planar)	70
Metatarsal (dorsal & planar)	70
Medial plantar fascia	200
Central plantar fascia	230
Lateral plantar fascia	180
Long plantar	70

Figure 14. Spring stiffness values for various ligaments (Muhammad Hanif Ramlee et.al. 2014)

3.2. Simulation results

In the simulations, we examined the displacements and the forces generated in the ligaments during the movements. These results are illustrated in the Figures below.

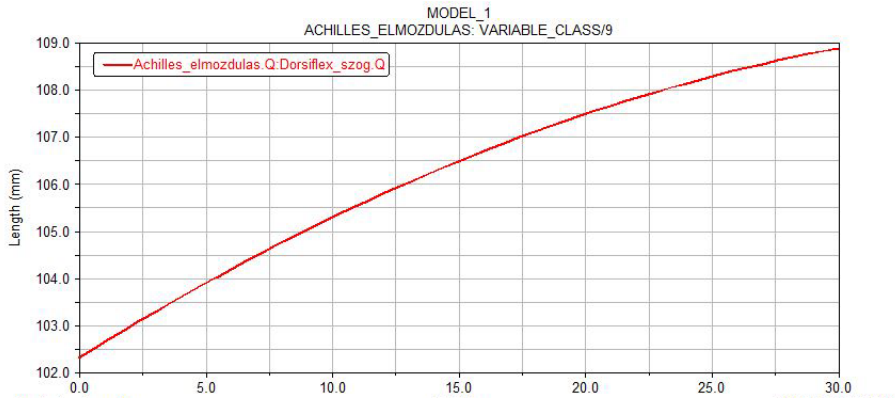


Figure 15. Change in Achilles tendon length during 30-degree Dorsiflexion

Figure 16. Az Change in Achilles tendon length during Plantarflexion with 40-degree rotation

Figure 15 shows the change in Achilles tendon length during 30° dorsiflexion, indicating a clear elongation as the foot moves upward. This reflects increased tensile loading in the tendon due to joint rotation. Figure 16 presents similar behaviour during 40° plantar flexion, where the tendon length changes in the opposite direction. The magnitude of displacement suggests asymmetry between the two motion types.

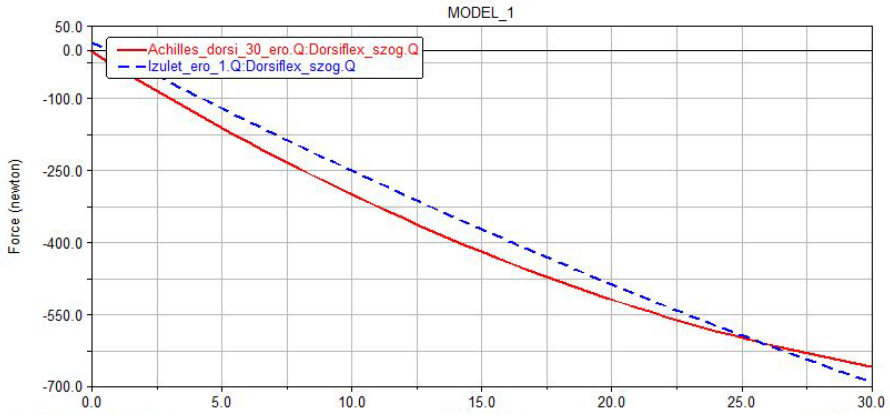


Figure 17. Changes in Achilles tendon and ligament forces during 30-degree Dorsiflexion

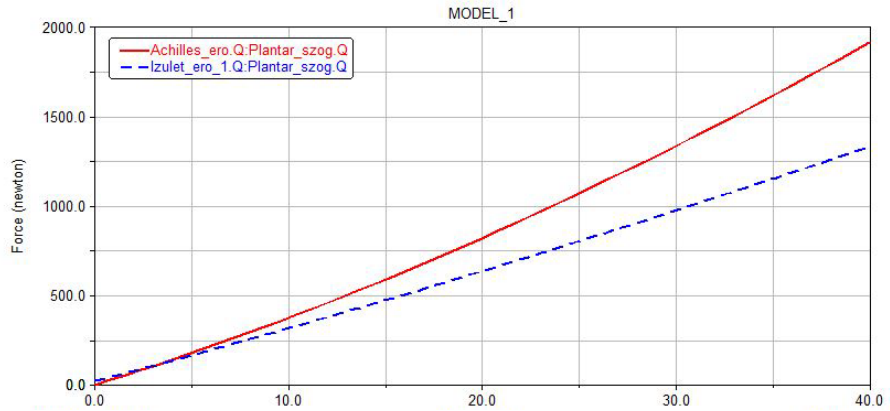


Figure 18. Changes in Achilles tendon and ligament forces during 40-degree Plantarflexion

Figure 17 highlights the variation in ligament and tendon forces during dorsiflexion. A progressive increase in force can be observed, corresponding to the stretching of ligament elements. Figure 18 demonstrates force distribution during plantar flexion, where different ligaments are activated. The comparison reveals that ligament loading patterns strongly depend on the direction of motion.

Overall, the figures confirm that the model effectively captures both displacement and force characteristics of ankle biomechanics.

4. Conclusions

The developed 3D ankle model and simulation methodology provide a robust and effective framework for analysing the complex kinematic and mechanical behaviours of the ankle joint. By integrating a scanned physical model with refined solid geometry and implementing multibody simulation techniques, the study achieved a realistic representation of joint motion. The application of kinematic constraints and the use of spring elements to model ligament behaviour allowed for the evaluation of internal forces and displacements under physiologically relevant conditions. The results demonstrate that even a simplified modelling approach can capture key biomechanical characteristics of the ankle, including the interaction between bones and soft tissues during dorsiflexion and plantar flexion. Furthermore, the model offers valuable insights into ligament loading patterns, which are critical for understanding injury mechanisms and rehabilitation processes. This approach can support future developments in clinical applications, such as orthotic design, surgical planning, and personalized treatment strategies, while also serving as a foundation for more advanced biomechanical simulations.

5. Acknowledgements

The research presented here was carried out as part of project 2018-1.3.1-VKE-2018-00013, funded by the NKFIH Fund.

References

[1] SW KONG (2020): Chronic Ankle Instability FAQs, Asia Medical Specialists.

[2] Muhammad Hanif Ramlee, Mohammed Rafiq Abdul Kadir, Malliga Raman Murali, Tunku Kamarul (2014): Finite element analysis of three commonly used external fixation devices for treating Type III pilon fractures, Medical Engineering & Physics Volume 36, Issue 10, pp. 1322-1330

[3] Hongyu Zhao, Sen Qiu, Zhelong Wang, Ning Yang, Jie Li and Jianjun Wang (2019): Applications of MEMS Gyroscope for Human Gait Analysis, 2. Gait characterization. 5-7. p.

[4] Ting-Cheng Chang, Xiao-Dong Zhang (2019): Kinematics and reliable analysis of decoupled parallel mechanism for ankle rehabilitation, 2. The motion of ankle joint. 204. p.Mondal D.P., S. Das, N and Jha (2009), Dry sliding wear behaviour of aluminum syntactic foam, Materials and Design, 30, 2563–2568

[5] Marta, G;Quental, C;(...); and Guerra-Pinto, F (2022): Multibody modelling of the foot for the biomechanical analysis of the ankle joint during running: A narrative review. PROCEEDINGS OF THE INSTITUTION OF MECHANICAL ENGINEERS PART K-JOURNAL OF MULTI-BODY DYNAMICS, DOI: 10.1177/14644193221090871

[6] Sancisi, N;Baldisserri, B;(...); and Leardini, A. (2014): One-degree-of-freedom spherical model for the passive motion of the human ankle joint MEDICAL & BIOLOGICAL ENGINEERING & COMPUTING, DOI: 10.1007/s11517-014-1137-y

[7] Jamwal, PK;Hussain, S;(...); and Quan, XS (2017): Musculoskeletal modelling of human ankle complex: Estimation of ankle joint moments. CLINICAL BIOMECHANICS, doi: 10.1016/j.clinbiomech.2017.03.010

Exploring artificial intelligence LLM adaptation strategies
a study of Retrieval-Augmented Generation (RAG) and Low-Rank Adaptation (LoRA)

Márk László LENKEI[1], Zsigmond Gábor SZALAY[2], István SZABÓ[3]
[1] *Independent IT advisor (*lmark995@gmail.com*)*
[2] *Department of Agricultural Digitalization and Advisory Services, Hungarian University of Life Sciences, Gödöllő, Hungary*
[3]*Institute of Technology, Hungarian University of Life Sciences, Gödöllő, Hungary*

Abstract

This paper examines two prominent strategies for adapting large language models (LLMs) to specialized tasks and knowledge domains: Retrieval-Augmented Generation (RAG) and Low-Rank Adaptation (LoRA), a form of Parameter-Efficient Fine-Tuning (PEFT). LLMs have come to play a central role in practical artificial intelligence applications, and solutions built upon them hold particular promise in fields where the professional knowledge base of workers is often incomplete or outdated, such as agriculture and various industrial domains. The study first explores the theoretical foundations and operating principles of each approach. LoRA addresses the prohibitive computational costs of full-parameter fine-tuning by freezing pretrained model weights and introducing trainable low-rank decomposition matrices, enabling efficient task-specific adaptation with minimal overhead and zero added inference latency. RAG augments parametric LLM knowledge with real-time retrieval from external databases, allowing models to access dynamic, domain-specific information without retraining, thereby reducing hallucination and improving factual accuracy. The paper further traces the evolution of RAG architectures from the foundational Naive RAG paradigm through Advanced and Modular RAG variants, each addressing limitations of its predecessor. Finally, real-world implementations of both methodologies are surveyed across diverse industries, including public administration, cybersecurity, agriculture, engineering, and healthcare, demonstrating their practical versatility and impact. Together, these adaptation strategies represent complementary solutions to the core challenge of deploying capable yet flexible LLMs in specialized, resource-constrained environments.

Keywords

Artificial Intelligence, Retrieval-Augmented Generation (RAG), Low-

Rank Adaptation (LoRa), Parameter-Efficient Fine-Tuning (PEFT), Large Language Model (LLM)

1. Introduction

Due to the rapid advancement of Artificial Intelligence and Natural Language Processing (NLP), the integration of Large Language Models (LLMs) has become an integral part of everyday workflows. The so-called Large Language Models (LLMs) have played a central role in the practical applications of artificial intelligence in recent times. Solutions built upon these—such as RAG techniques—can serve as excellent tools in the development of various AI-based expert systems, and their impact can be of paramount importance in fields like agriculture or industrial applications (Tarr-Szabó, 2025). This is because the professional preparedness of workers in agriculture and technical fields is often incomplete or is not based on the most up-to-date knowledge base. Although Transformer-based architectures possess outstanding generative capabilities, their application faces significant limitations in tasks requiring specific, up-to-date information or niche domain expertise. The primary source of these obstacles is the models' static training datasets and the fixed state of their parameters. The full-parameter fine-tuning of Large Language Models (LLMs) often encounters hardware constraints due to significant computational resource requirements and high memory costs. For instance, the full fine-tuning of a 7-billion-parameter LLaMA model using the conventional AdamW optimizer can require up to 102 GB of GPU memory (Lv et al., 2024). To improve memory optimization and cost-efficiency, several relevant methodologies have emerged in the literature, among which the present work highlights the following:

1. LoRA (Low-Rank Adaptation): A specific type of Parameter-Efficient Fine-Tuning (PEFT) that implements adaptation by utilizing low-rank approximations of the pre-trained model's weight matrices. This method enables the task-specific tuning of models with a minimal number of additional parameters, thereby optimizing resource consumption (Hu et al., 2021).
2. RAG (Retrieval-Augmented Generation): A hybrid NLP architecture that combines the parametric memory (internal weights) of large language models with real-time information retrieval from external, non-parametric databases. This approach allows for the integration of dynamic, domain-specific knowledge into the generative process without the need for costly model retraining, thereby minimizing hallucination rates and ensuring the factual accuracy of outputs (Lewis et al., 2020).

2. Low-Rank Adaptation operating principles

The theoretical foundations of the LoRA (Low-Rank Adaptation) method were established by Li et al. (2018) and Aghajanyan et al. (2020). These studies demonstrated that large, heavily overparameterized pretrained language models possess a low intrinsic dimensionality. This implies that such models can learn efficiently even when their parameters are projected onto a significantly smaller subspace. Based on this insight, the hypothesis emerged that weight updates performed during fine-tuning for downstream tasks also exhibit a low intrinsic rank structure. LoRA leverages this observation by approximating weight updates using a low-rank decomposition rather than modifying the full weight matrix (Hu et al., 2021).

The core mechanism of LoRA operates by freezing the weights of pretrained dense layers and introducing a parallel, trainable low-rank pathway. For any pretrained weight matrix $W_0 \in R^{d \times k}$, the standard gradient-based update ΔW is structurally constrained via a low-rank decomposition:

$$W_0+\Delta W = W_0+BA$$

where $B \in R^{d \times r}$ and $A \in R^{r \times k}$. The hyperparameter r denotes the rank of the decomposition and is intentionally chosen such that $r \ll min(d,k)$. (Hu et al., 2021).

During stochastic optimization, the base matrix W_0 remains completely frozen and receives no gradient updates, which significantly reduces the memory requirements associated with optimizer states (Hu et al., 2021). The adaptation mechanism exclusively optimizes the parameters within matrices A and B. During the forward pass, the input vector x is multiplied both by the frozen weights and the low-rank matrices. The resulting activation vectors are summed element-wise, yielding the hidden representation h:

$$h=W_0x+\Delta W_x=W_0x+BAx$$

The functional stability of the LoRA mechanism strongly depends on its initialization scheme. To ensure that the model initially behaves identically to the pretrained base model, matrix A is initialized with a random Gaussian distribution, while matrix B is initialized to zero. Consequently, the initial update satisfies $\Delta W=BA=0$ (Hu et al., 2021).

This parallel architecture provides a key practical advantage. In contrast to sequential approaches such as the adapter modules proposed by Houlsby et al. (2019) and Pfeiffer et al. (2020) which increase network depth and inference latency, LoRA avoids these drawbacks. The trained matrices A and B can be explicitly merged with the original weights prior to deployment: $W_{merged}=W_0+BA$ This linear fusion guarantees zero additional latency, as the deployed model

performs the standard computation $h=W_{merged}x$ during inference (Hu et al., 2021).

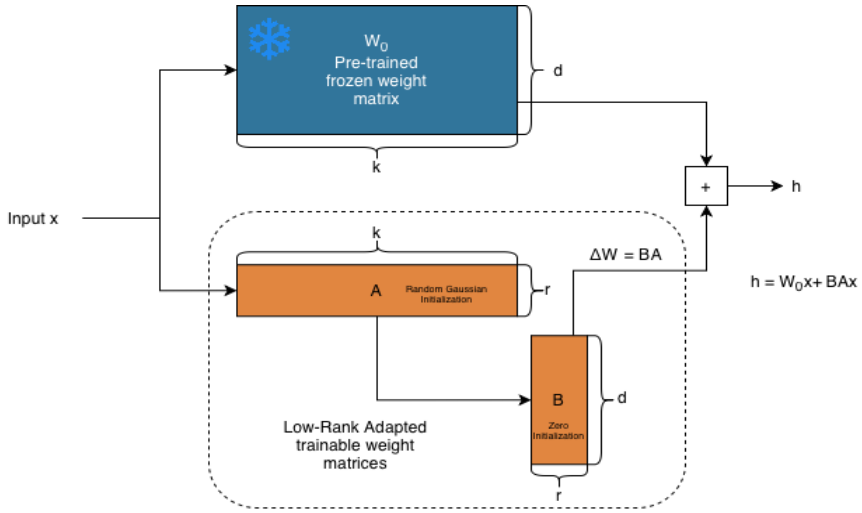


Figure 1: Architecture of Low-Rank Adaptation (LoRA). During fine-tuning, the pre-trained weight matrix (W_0) remains frozen. Task-specific updates are learned through a parallel trainable pathway consisting of two low-rank matrices, A and B. Matrix A compresses the input dimension (k) to a much smaller rank (r), and Matrix B expands it back to the output dimension (d). Initializing B with zeros ensures the adaptation module behaves as an identity function at the start of training. Architecture based on Hu et al. (2021).

3. Retrieval-Augmented Generation operating principles

Retrieval-Augmented Generation (RAG) represents a hybrid approach to natural language processing that endows pre-trained, parametric language models with a non-parametric memory. This architecture addresses the limitations of relying solely on the implicit knowledge stored within a model's weights by introducing a differentiable access mechanism to an explicit, external knowledge source (Lewis et al., 2020).

The RAG mechanism relies on two primary components that work in tandem:

1. The Retriever (Non-Parametric Memory): Denoted as $p_\eta(z|x)$, this component returns a distribution over text passages z given an input query x. It typically utilizes a bi-encoder architecture, where a dense embedding model maps both the documents and the query into a shared vector space. By employing Maximum Inner Product Search (MIPS) or similar vector similarity metrics, the retriever efficiently identifies the top-K documents with the highest relevance to the input (Lewis et al., 2020).

2. The Generator (Parametric Memory): Modeled as $p_\theta(y_i|x,z,y_{1:i-1})$, this component is a sequence-to-sequence transformer. It generates the target sequence y by conditioning on a concatenation of the original input x and the retrieved context z (Lewis et al., 2020).

To integrate retrieval and generation end-to-end, RAG treats the retrieved document as a latent variable. The model marginalizes over these latent documents to produce a distribution over the generated text via two primary formulations:

1. RAG-Sequence: This model assumes that the same retrieved document is used to predict every target token in a sequence. It treats the document as a single latent variable marginalized via a top-K approximation to calculate the overall sequence probability (Lewis et al., 2020).
2. RAG-Token: In this approach, the model can draw a different latent document for each individual target token. This allows the generator to dynamically synthesize and combine content from multiple distinct documents while producing a single, cohesive answer (Lewis et al., 2020).

A crucial aspect of the RAG mechanism is its joint training procedure. The query encoder of the retriever and the generator are fine-tuned end-to-end to minimize the negative marginal log-likelihood of the target sequence. This joint learning occurs without direct supervision of which documents should be retrieved. To maintain computational efficiency, the document embeddings and the external index are typically kept fixed during the fine-tuning process (Lewis et al., 2020).

With the widespread adoption of powerful embedding models for retrieval such as BGE-M3 (Chen et al., 2024), Gemini Embedding (Lee et al., 2025) alongside advanced Large Language Models (LLMs) like ChatGPT, Gemini or Claude with strong in-context learning capabilities, the focus has shifted toward architectures that combine efficient semantic search with context-aware generation, giving rise to the Naive RAG paradigm. This approach follows a basic "Retrieve-Read" framework consisting of simple indexing, retrieval, and generation steps. (Gao et al., 2023)

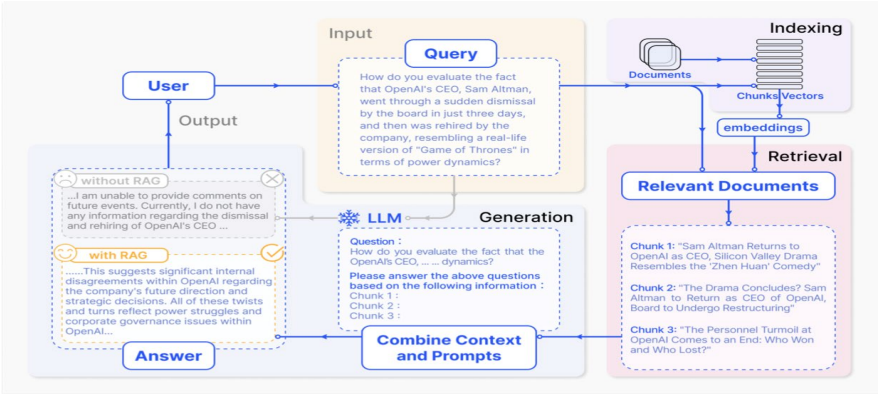


Figure 2. A representative instance of the RAG process applied to question answering. It mainly consists of 3 steps. 1) Indexing. Documents are split into chunks, encoded into vectors, and stored in a vector database. 2) Retrieval. Retrieve the Top k chunks most relevant to the question based on semantic similarity. 3) Generation. Input the original question and the retrieved chunks together into LLM to generate the final answer. From *Retrieval-augmented generation for large language models: A survey* by Y. Gao et al., 2023, arXiv (https://doi.org/10.48550/arXiv.2312.10997)

However, Naive RAG frequently encounters limitations, such as struggling with retrieval precision and recall, retrieving misaligned chunks, and generating outputs that suffer from information overload or disjointed integration of context. To overcome these hurdles, the technology evolved into Advanced RAG, which employs targeted pre-retrieval and post-retrieval strategies. Pre-retrieval optimizations enhance the quality of the retrieved content through better data granularity, sliding window chunking, query rewriting, and query expansion. Post-retrieval processes refine the results by reranking the most relevant documents to the edges of the prompt and compressing the context to prevent the LLM from losing focus amidst irrelevant noise. (Gao et al., 2023)

Furthermore, the field has also introduced Modular RAG, which offers a highly adaptable and versatile architecture that moves beyond simple sequential retrieval. Modular RAG introduces specific functional modules like routing mechanisms, search modules, and memory pools and employs sophisticated interaction patterns such as iterative, recursive, and adaptive retrieval, which allow the model to actively judge when and what to retrieve (Gao et al., 2023).

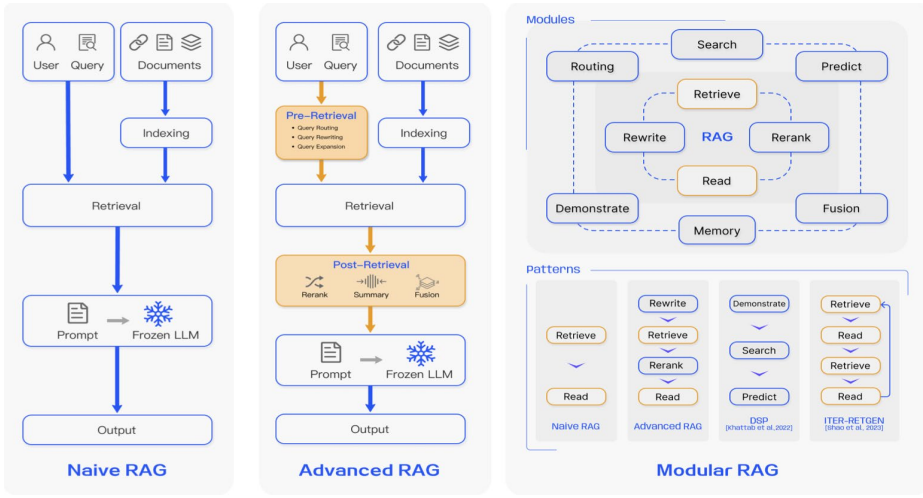


Figure 3. Comparison between the three paradigms of RAG. (Left) Naive RAG mainly consists of three parts: indexing, retrieval and generation. (Middle) Advanced RAG proposes multiple optimization strategies around pre-retrieval and post-retrieval, with a process similar to the Naive RAG, still following a chain-like structure. (Right) Modular RAG inherits and develops from the previous paradigm, showcasing greater flexibility overall. This is evident in the introduction of multiple specific functional modules and the replacement of existing modules. The overall process is not limited to sequential retrieval and generation; it includes methods such as iterative and adaptive retrieval. From Retrieval-augmented generation for large language models: A survey by Y. Gao et al., 2023, arXiv (https://doi.org/10.48550/arXiv.2312.10997)

4. Real-life use cases

This section examines real-world implementations of Low-Rank Adaptation (LoRa) and Retrieval-Augmented Generation (RAG) to assess their versatility and impact within the contemporary AI landscape. The practical utility of Low-Rank Adaptation (LoRA) is demonstrated through its integration into diverse, real-world deployment scenarios, including:

1. **Daily NLP Operations:** LoRA is used to fine-tune large language models for emotion recognition in conversations, text classification, and role recognition. It also powers document-level relation extraction, machine translation, and general text understanding (Mao et al., 2024).
2. **Software Development:** Developers use LoRA-adapted models for Just-In-Time defect prediction during code changes, code summarization, and detecting cloned code. LoRa such as RepairLLaMA is also applied to automated program repair. Besides, it can be also used in database management tasks like query rewriting and index tuning(Mao et al., 2024).
3. **Healthcare & Finance:** In the medical field, LoRA fine-tunes models for clinical dialogue summarization, assertion detection, and answering

medical questions. In the financial sector, it is used for analyzing financial news and sentiment classification (Mao et al., 2024).

4. **Digital Art & Media Generation:** LoRA is integrated into diffusion models to handle text-to-image generation, image editing, and transferring artistic styles onto images. It is also used in the generation of videos and 3D content (Mao et al., 2024).
5. **Autonomous AI Agents:** LoRA modules can be used to assign highly complex roles to autonomous AI agents, act as specific tools the agents can use, or even serve as memory modules for continual learning (Mao et al., 2024).

Several specialized Retrieval-Augmented Generation (RAG) systems have been developed by Hasan et al. (2026) to address industry-specific challenges:

1. **Kankaanpää City AI:** This RAG-based implementation improves public access to government data by indexing more than 1,000 administrative PDFs from 2023 and 2024. Documents are transformed into embeddings using text-embedding-ada-002 and stored in a FAISS vector database. Integration with gpt-4o-mini allows the system to deliver accurate, context-specific insights from complex policy records, streamlining the way users retrieve and interact with city information (Hasan et al., 2026).
2. **Disarm RAG:** This implementation provides a secure environment for cybersecurity research and real-time forensic investigation. By deploying LLaMA 2-uncensored within the sovereign infrastructure of the Finnish IT Center for Science (CSC), the system ensures complete data sovereignty and privacy. It integrates the DISARM Framework to bridge the gap between offensive tactics—such as phishing and deep-fake creation—and defensive protocols like misinformation control. The RAG-driven interface allows investigators to query both attack vectors and mitigation strategies, facilitating a comprehensive understanding of modern threats like multi-factor authentication (MFA) bypasses and deep-fake-based phishing (Hasan et al., 2026).
3. **AgriHubi AI Assist:** This RAG implementation serves as a digital bridge for the agricultural sector, transforming 200+ Finnish-language PDFs into an accessible knowledge base through multilingual OCR and FAISS vectorization. At its core, the system utilizes the Poro-34B model—specifically tuned for the Finnish language—to deliver precise guidance on soil conservation and sustainable practices. The deployment features a user-centric Streamlit chat interface supported by an SQLite backend for interaction logging and continuous performance improvement. By streamlining the retrieval of complex policy and technical data, AgriHubi AI Assist enhances the decision-making capabilities of agricultural professionals (Hasan et al., 2026).

4. **FEMMa Oracle:** This RAG-based implementation streamlines the exploration of electrified mobile machinery research by transforming specialized PDF documentation into a searchable, structured knowledge base. Utilizing a combination of text-embedding-3-large for precise vectorization and GPT-4o for natural language generation, the system provides a robust interface for querying validated engineering records. This setup eliminates the friction typically associated with manual document review, allowing for more agile and accurate retrieval of technical specifications and project outcomes (Hasan et al., 2026).
5. **Assist Doctor:** This implementation serves as a specialized knowledge bridge for aneurysm diagnostics, facilitating evidence-based decision-making across neurosurgical and radiological disciplines. The system transforms a corpus of clinical data and scholarly research into a searchable vector space, utilizing an advanced embedding pipeline to ensure precise information retrieval. Hosted via a Streamlit interface and powered by GPT-4, the tool allows practitioners to query intricate risk models and treatment protocols in real-time. By streamlining the synthesis of clinical literature, Assist Doctor enhances the accuracy and efficiency of diagnostic workflows in high-stakes vascular care (Hasan et al., 2026).

6. Conclusions

The rapid evolution of large language models has created an urgent need for practical adaptation strategies that balance performance, efficiency, and flexibility. This paper has examined two methodologies addressing this challenge: Low-Rank Adaptation (LoRA) and Retrieval-Augmented Generation (RAG). LoRA has demonstrated its value as a computationally efficient fine-tuning approach, enabling task-specific model customization across a remarkably broad range of applications. RAG, meanwhile, has proven indispensable for knowledge-intensive applications where factual accuracy and up-to-date information are paramount. Crucially, these two methodologies are not mutually exclusive but rather complementary. LoRA excels at embedding domain-specific reasoning directly into model parameters, while RAG provides dynamic access to external knowledge that static training cannot capture. Hybrid architectures combining both approaches are already an active area of scholarly inquiry, with works such as P-RAG (Lyu et al., 2025), which integrates LoRA with chain-of-thought reasoning for biomedical question answering, and AutoRAG-LoRA (Dwivedi & Mishra, 2025), which deploys lightweight LoRA adapters to dynamically correct hallucinations in RAG systems, pointing toward a promising convergence of the two paradigms. As the field continues to mature, the integration of these complementary strategies will remain central to realizing the full potential of large language models in practical, real-world settings.

References

[1] Tarr B. – Szabó I. (2025) Egyszerűség vagy komplexitás? Prediktiv mesterséges intelligencia modellek összehasonlítása a mezőgazdaságban Mezőgazdasági Technika 66.évf. 7.szám, ISSN 0026-1890

[2] Lv, K., Yang, Y., Liu, T., Gao, Q., Guo, Q., & Qiu, X. (2024). *Full parameter fine-tuning for large language models with limited resources*. arXiv. https://doi.org/10.48550/arXiv.2306.09782

[3] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models*. arXiv. https://doi.org/10.48550/arXiv.2106.09685

[4] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. arXiv. https://doi.org/10.48550/arXiv.2005.11401

[5] Li, C., Farkhoor, H., Liu, R., & Yosinski, J. (2018). *Measuring the intrinsic dimension of objective landscapes. arXiv*. https://doi.org/10.48550/arXiv.1804.08838

[6] Aghajanyan, A., Zettlemoyer, L., & Gupta, S. (2020). *Intrinsic dimensionality explains the effectiveness of language model fine-tuning. arXiv*. https://doi.org/10.48550/arXiv.2012.13255

[7] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. (2019). *Parameter-efficient transfer learning for NLP. arXiv*. https://doi.org/10.48550/arXiv.1902.00751

[8] Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., & Gurevych, I. (2020). *AdapterFusion: Non-destructive task composition for transfer learning. arXiv*. https://doi.org/10.48550/arXiv.2005.00247

[9] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., et al. (2023). *Retrieval-augmented generation for large language models: A survey*. arXiv. https://doi.org/10.48550/arXiv.2312.10997

[10] Lee, J., Chen, F., Dua, S., Cer, D., Shanbhogue, M., Naim, I., Ábrego, G. H., Li, Z., Chen, K., Vera, H. S., Ren, X., Zhang, S., Salz, D., Boratko, M., Han, J., Chen, B., Huang, S., Rao, V., Suganthan, P., . . . Duerig, T. (2025). *Gemini embedding: Generalizable embeddings from Gemini*. arXiv. https://doi.org/10.48550/arXiv.2503.07891

[11] Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). *M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation*. arXiv. https://doi.org/10.48550/arXiv.2402.03216

[12] Hasan, M. T., Waseem, M., Kemell, K.-K., Khan, A. A., Saari, M., & Abrahamsson, P. (2026). Engineering RAG systems for real-world applications: Design, development, and evaluation. In *Proceedings of the 51st Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2025:*

Lecture Notes in Computer Science (Vol. 16082, pp. 143–158). Springer. https://doi.org/10.1007/978-3-032-04200-2_10

[13] Mao, Y., Ge, Y., Fan, Y., Xu, W., Mi, Y., Hu, Z., & Gao, Y. (2024). A survey on LoRA of large language models. *Frontiers of Computer Science*. Advance online publication. https://doi.org/10.1007/s11704-024-40663-9

[14] Lyu, X., Lyu, G., Yan, Z., & Jiang, Y. (2025). P-RAG: Prompt-Enhanced Parametric RAG with LoRA and Selective CoT for Biomedical and Multi-Hop QA. *Proceedings of International Conference on Computing and Data Science Symposium: Application of Machine Learning in Engineering*. https://doi.org/10.54254/2755-2721/2025.AST28253

[15] Dwivedi, K., & Mishra, P. P. (2025). *AutoRAG-LoRA: Hallucination-triggered knowledge retuning via lightweight adapters* (arXiv:2507.10586). arXiv. https://doi.org/10.48550/arXiv.2507.10586

Numerical comparison of flat plate absorber configurations for solar air heaters using ansys fluent

Abraham Kiprop JUMA [1], Zoltán KURJÁK [2] Janos BEKE [2]

[1] *Doctoral Program of Mechanical Engineering, Hungarian University of Agriculture and Life Sciences,* MATE Gödöllő

[2] *Department of Vehicle Technology, Institute of Technology, Hungarian University of Agriculture and Life Sciences,* MATE Gödöllő

Abstract

Solar air heaters are pivotal components of indirect solar dryers. The thermal and hydraulic performance of these heaters directly influences energy consumption, air distribution, and drying efficiency. Conventional flat plate solar air heaters dominate the drying field due to their simplicity and cost-effectiveness. However, flat plate air heaters are characterized by low thermal efficiency resulting from weak convective heat transfer between the flowing air and the absorber surface. This study undertakes a numerical comparison of three flat plate absorber configurations: a plain flat plate, a finned flat plate, and a 60° V-grooved flat plate. CFD simulations were performed using ANSYS Fluent under similar operating conditions, geometric dimensions, and turbulence models. The analysis focuses on outlet air temperature, velocity distribution, and pressure drop. The results show that surface modification of flat plate absorbers leads to significant enhancement in heat transfer. The plain flat plate configuration records the lowest outlet air temperature and the lowest pressure drop. The finned configuration provides a balance between thermal and hydraulic performance, making it suitable for low- to medium-temperature solar drying applications. The V-grooved absorber plate yields the highest outlet air temperature, but at the cost of a significantly higher pressure drop. Overall, the findings highlight the thermo-hydraulic trade-offs among different flat plate absorber configurations and provide design guidance for selecting an appropriate absorber based on specific drying requirements.

Keywords

Solar air heater, Flat plate absorber, Thermo-hydraulic performance, Computational Fluid Dynamics (CFD), solar drying

1. Introduction

Solar drying is one of the most widely studied renewable energy technologies for post-harvest preservation of agricultural products. It offers a low-cost, clean,

and sustainable alternative to fossil-fuel-based and open-sun drying systems (Devan et al., 2020). This technology is used to dehydrate agricultural produce to safe storage moisture levels, thereby minimizing post-harvest losses, extending shelf life, and enhancing product quality (Baidhe et al., 2024). The benefits of solar drying are particularly significant in developing countries, where agriculture forms the backbone of the economy, yet post-harvest losses remain high due to limited access to adequate preservation resources.

Indirect solar dryers have gained considerable attention because they provide improved drying control and superior product quality, as the produce is not directly exposed to solar radiation, unlike in direct solar dryers (Deepak & Behura 2023). These dryers typically comprise a solar air heater, an airflow system, and a drying chamber. The solar air heater plays a central role, as it determines the drying air temperature, airflow uniformity, and flow rate. Its performance is a key factor in energy efficiency, drying time, and the overall effectiveness of the drying process. Consequently, the thermal and hydraulic performance of solar air heaters has become a major focus of research in solar thermal applications.

Solar air heaters can be broadly classified into three main types: flat plate, evacuated tube, and concentrating collectors (Kalair et al., 2022). Although evacuated tube and concentrating collectors offer superior thermal performance, flat plate solar air heaters remain the most widely used in solar drying applications due to their simple design, low cost, ease of fabrication and installation, and suitability for low- to medium-temperature operation (Vengadesan & Senthil, 2020). Flat plate solar air heaters typically operate within a temperature range suitable for drying a wide variety of agricultural products, including vegetables, fruits, grains, herbs, and spices.

Despite their widespread use, conventional flat plate solar air heaters exhibit relatively low thermal efficiency. This limitation is primarily due to poor convective heat transfer between the flowing air and the absorber plate. The formation of a stable thermal boundary layer along the smooth absorber surface restricts heat transfer, resulting in substantial heat losses. Consequently, larger collector surface areas and longer drying times are often required to achieve the desired drying conditions.

To overcome these limitations, numerous studies have proposed modifications to the absorber plates of flat plate solar air heaters. Widely investigated modifications include the use of baffles, fins, corrugations, ribs, and grooved surfaces (El-Sebaey et al., 2024). These modifications aim to increase the effective heat transfer surface area and induce turbulence within the airflow, thereby enhancing flow mixing and improving the uniformity of drying air characteristics. While such techniques significantly improve thermal performance, they also increase flow resistance, leading to higher pressure drops and additional energy consumption for air circulation (Desisa & Shekata, 2020).

Finned flat plate absorbers have been extensively studied in solar air heater applications. This approach involves adding fins to the absorber plate to increase

the heat transfer surface area and promote turbulence, thereby enhancing convective heat transfer. Experimental and numerical studies have reported significant improvements in thermal efficiency and outlet air temperature for finned absorber plates compared to plain flat plate configurations. Kansara, Pathak, & Patel (2021) reported that the outlet air temperature for finned collectors was 8.19% higher than that of smooth plates. Moreover, finned absorber plates generally impose relatively low pressure drops, making them suitable for small- to medium-scale solar drying systems where improved thermal performance is desired without substantially increasing energy consumption.

Another commonly applied modification to flat plate absorbers is the use of grooved surfaces, particularly the 60° V-grooved configuration. This design promotes repeated flow separation and reattachment, generating strong turbulence and air recirculation within the grooves. Compared to plain and finned absorber plates, V-grooved surfaces provide enhanced flow mixing and convective heat transfer. Studies have shown that V-grooved solar air heaters achieve higher outlet air temperatures than other flat plate configurations (Oliveira & Dutra, 2023). However, the increased turbulence is accompanied by significant pressure drops, which can limit their applicability in low- to medium-temperature solar drying systems with constrained blower power (Abd, Alomar, & Salih, 2022).

Although numerous studies have investigated various flat plate absorber configurations, most have primarily focused on thermal efficiency and outlet air temperature. While these parameters are important, velocity distribution and pressure drop are equally critical in solar drying applications, as they directly influence airflow uniformity, drying quality, and energy consumption. Furthermore, limited research has conducted direct numerical comparisons of plain, finned, and V-grooved configurations under identical operating conditions, geometric parameters, and turbulence modeling approaches.

In recent years, computational fluid dynamics (CFD) has emerged as a powerful tool for analyzing heat transfer and fluid flow in solar air heaters. CFD enables detailed visualization of temperature, velocity, and pressure fields, which are often difficult to obtain experimentally. When applied using consistent modeling assumptions and boundary conditions, CFD provides a robust framework for systematic comparison and design optimization of different solar air heater configurations.

The present study employs ANSYS Fluent to systematically compare three flat plate solar air heater configurations: plain flat plate, finned flat plate, and V-grooved flat plate, under identical geometric and operating conditions. The analysis simultaneously evaluates pressure drop, outlet air temperature, and velocity distribution, which are key parameters in assessing the performance of solar dryers. The novelty of this work lies in the integrated numerical evaluation of both thermal and hydraulic behavior of flat plate solar air heater

configurations, with particular emphasis on practical design trade-offs for low- to medium-temperature solar drying applications.

2. Numerical methodology

2.1. Geometry Description

This study investigated three configurations of flat-plate solar collectors: a finned flat plate, a plain flat plate, and a 60° V-grooved absorber plate. To ensure a fair comparison, all configurations were designed with identical external dimensions. The length, width, and height of the air channels were 150 mm × 100 mm × 30 mm for all prototypes, with only the absorber surface geometry varied, as shown in Figure 1. All absorber plates were made of aluminum with a thickness of 2 mm.

The plain flat plate, commonly used in small-scale solar dryers, served as the baseline for comparison. The finned absorber plate featured eight transverse fins (four fins extending halfway from the edge to the center of the plate), each with a height of 7 mm and a width of 9 mm, evenly distributed along the airflow direction. The 60° V-grooved configuration, widely reported in the literature for its enhanced thermal performance in air heaters, included nine symmetric V-grooves uniformly positioned along the airflow path. Each groove had a depth of 1 mm.

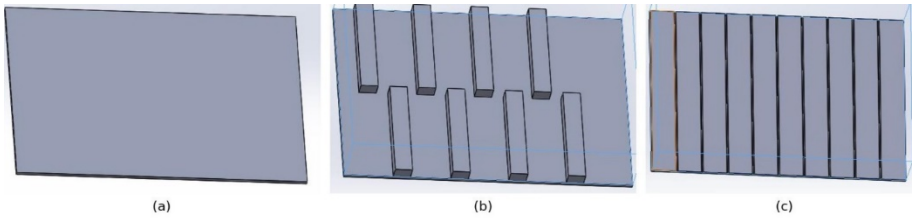


Figure 1. Schematic comparison of flat absorber plate configurations: (a) plain absorber plate (b) finned absorber plate, and (c) V-grooved absorber plate.

2.2. Mesh generation

A structured mesh was employed in ANSYS Fluent to discretize the computational domain, which included the absorber plate (solid domain) and the air channel (fluid domain). Tetrahedral elements were used for the finned and V-grooved absorber geometries, while hexahedral elements were applied to the plain flat plate to enhance numerical accuracy. The mesh was refined near the absorber surfaces to capture flow boundary layers and thermal gradients accurately. Five inflation layers were generated along the fluid–solid interfaces with a growth rate of 1.2 to ensure adequate resolution of near-wall turbulence effects.

Mesh quality was evaluated using skewness, with a maximum skewness maintained below 0.90. A grid independence study was performed by progressively refining the mesh until variations in pressure drop and outlet air temperature were below 1%, ensuring reliable numerical results.

2.3. Turbulence Model and the Governing Equations

The airflow was assumed to be incompressible, steady, and turbulent. The other assumptions in the study are radiation loss is negligible, there is uniform heat flux applied in the absorber surface, and there is no internal heat generation in air. The turbulence effects were modeled based on the k–ε turbulence, which the existing literature has reported to provide better accuracy taking into account the computational cost.

The simulation has been carried based on the continuity equation, the energy equation, and the three-dimensional Navier-Stokes equations.

Continuity Equation

$$\partial\rho/\partial t + \nabla\cdot(\rho U) = 0$$

Where:

- ρ is the fluid density (kg m^{-3})
- U is the velocity vector (m s^{-1})
- t is time (s)
- $\nabla\cdot$ denotes the divergence operator

Momentum Equations (the dimensional Navier–Stokes equation)

$$\partial(\rho U)/\partial t + \nabla\cdot(\rho U \otimes U) = -\nabla p + \nabla\cdot\tau + S_M$$

Where:

- p is the static pressure (Pa)
- τ is the viscous stress tensor (Pa)
- S_M represents momentum source terms
- $\otimes$ denotes the tensor product

Stress tensor definition:

$\tau = \mu [\nabla U + (\nabla U)^T - (2/3) \delta \nabla\cdot U]$

The stress tensor defines the viscous stress in a Newtonian fluid

Where:

- μ is the dynamic viscosity (Pa·s)
- $(\nabla U)T$ is the transpose of the velocity gradient tensor
- Δ is the Kronecker delta

Total Energy Equation

$$\partial(\rho h_{tot})/\partial t - \partial p/\partial t + \nabla\cdot(\rho\, U\, h_{tot}) = \nabla\cdot(\lambda\, \nabla T) + \nabla\cdot(U \cdot \tau) + U \cdot S_M + S_E$$

Where:

- T is temperature (K)
- λ is the thermal conductivity (W m^{-1} K^{-1})
- S_E is an external energy source term

Total enthalpy definition:
$h_{tot} = h(T, p) + 1/2\ |U|^2$

Where:

- h is the static enthalpy (J kg^{-1})
- $1/2|U|^2$ represents kinetic energy per unit mass.

2.4. Boundary Conditions

For the flat plate air heaters modeled in this study, the inlet air velocity was assumed uniform and corresponded to a typical mass flow rate suitable for drying applications. The inlet air velocity was set at 0.5 m/s, with an ambient temperature of 300 K. The outlet boundary condition was defined as the prevailing atmospheric pressure (0 bar). Adiabatic conditions were assumed for all other boundaries. A constant heat flux of 800 W/m^2 was applied to the absorber plate, representing solar irradiation.

2.5. Solver Setup

The computational fluid dynamics simulations were performed using ANSYS Fluent with a pressure-based solver. The SIMPLE algorithm was used for pressure–velocity coupling. Second-order upwind discretization was applied for momentum and energy equations. Convergence was assumed when both the outlet air temperature and pressure drop values stabilized, and when the residuals for momentum, continuity, energy, and turbulence equations fell below 10^{-6}.

3. Results and Discussion

3.1. Temperature Distribution

The three absorber configurations exhibited distinct temperature profiles, as illustrated in the temperature contours in Figure 2. The plain flat absorber plate showed a relatively uniform temperature distribution but recorded the lowest outlet air temperature of 310K due to limited convective heat transfer. The smooth surface of the plain plate promotes the formation of a stable thermal boundary layer, which restricts heat exchange between the airflow and the absorber.

The finned configuration achieved moderately higher outlet temperatures compared to the plain plate. The addition of fins increased the heat transfer surface area and disrupted the thermal boundary layer on the smooth plate, enhancing convective heat transfer. The outlet air temperature for the finned plate was 312.5K.

The 60° V-grooved flat plate exhibited the highest fluid temperatures along the collector. The grooves generated strong turbulence and air recirculation, effectively disrupting the thermal boundary layer and promoting robust convective heat transfer. Among the three configurations, the V-grooved absorber achieved the highest outlet temperatures, 313K. These results indicate that V-grooved surfaces are particularly effective for drying applications that require elevated temperatures.

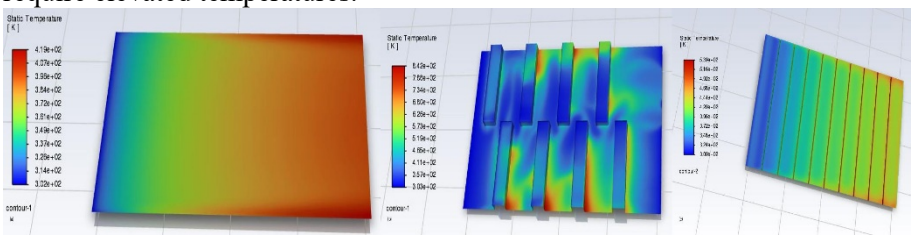


Figure 2. Numerical simulation results comparing the thermal performance of three absorber configurations: a, Plain, (b) fin, and (c) Groove

3.2. Velocity Distribution

The simulations also revealed distinct velocity profiles among the three flat plate configurations, summarized in Table 1. The plain flat plate exhibited relatively uniform and smooth airflow, as the working fluid encountered minimal disturbance. The finned configuration introduced moderate disturbances, enhancing flow mixing while maintaining a relatively uniform velocity profile suitable for solar drying.

In contrast, the V-grooved configuration produced significant velocity fluctuations due to recirculation zones formed within the grooves. While these grooves enhance convective heat transfer, the resulting non-uniform velocity

distribution can hinder uniform drying if not properly managed. However, the inlet and outlet velocity were nearly the same for both configurations as the simulation was pressure based.

Table 1. Velocity distribution among different absorber configurations.

Plate configuration	Inlet velocity (m/s)	Outlet velocity (m/s)
Plain	0.5	0.5
Finned	0.5	0.5
V-grooved	0.5	0.5

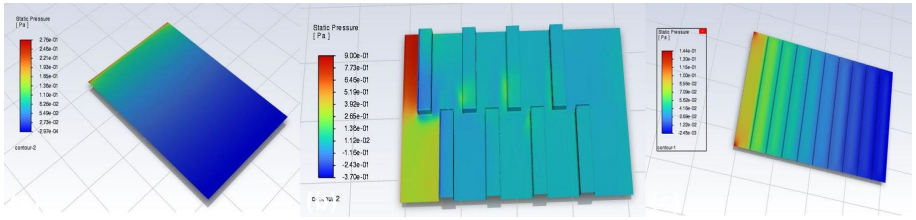
3.1. Pressure Drop

Pressure drop is a significant variable in solar drying as it directly determines the blower power required, which affects the overall energy efficiency of the dryer. Table 2. denotes the pressure drop profile in the three absorber plate configurations.

Table 2. Numerical evaluation of the inlet pressure required for different absorber configurations.

Plate configuration	Inlet Pressure	Outlet Pressure
Plain	0.12Pa	0Pas
Finned	0.36Pa	0Pa
V-grooved	3.4Pa	0pPa

The plain flat-plate absorber recorded the lowest pressure drop due to its smooth surface and minimal flow obstruction. A moderate pressure drop was observed for the finned plate configuration compared to the plain plate. The presence of fins introduced localized turbulence and additional frictional surface area, which increased the flow resistance. The V-grooved configuration exhibited the highest pressure drop among the three configurations. Repeated flow separation, recirculation, and reattachment within the grooves significantly increased the resistance to airflow. Although the V-grooved configuration offers superior thermal performance, the associated pressure losses limit its applicability in solar drying systems, which are typically energy-constrained. Figure 3 provide a comparative overview of the pressure drop among the three absorber configurations.



(a) (b) (c)

Figure 3. Numerical simulation results comparing the pressure profile of the three absorber configurations: a, Plain, (b) fin, and (c) Groove

3.2 Thermo-Hydraulic Performance

Thermal and hydraulic performance are critical factors when evaluating the effectiveness of flat-plate absorbers. Each configuration involves inherent trade-offs, as summarized in Table 3. Plain flat-plate absorbers exhibit minimal pressure drop but suffer from poor thermal performance. In contrast, the 60° V-grooved absorber provides superior thermal performance but incurs a significantly higher pressure drop.

The finned absorber offers a balanced compromise between thermal enhancement and hydraulic penalty. It achieves a substantial increase in outlet air temperature while maintaining a relatively moderate pressure drop. In the context of low- to medium-temperature solar drying applications, the finned flat-plate configuration represents a practical solution. Such applications require uniform airflow, improved thermal performance, and low auxiliary energy consumption—conditions that are not adequately met by either plain flat-plate absorbers or highly obstructive V-grooved configurations.

Table 3. A qualitative comparison of thermo-hydraulic performance of different absorber configurations.

Absorber Configuration	**Thermal Performance**	**Pressure Drop**	**Practical Applications**
Plain Flat Plate	Low	Minimal	Low-cost, Small scale
Finned Plate	High	Moderate	General solar drying
V-Grooved	Maximum	High	High-temperature

4. Conlcusion

The present study conducted an in-depth numerical investigation of three flat-plate absorber configurations: (i) plain flat plate, (ii) finned flat plate, and (iii)

60° V-grooved flat plate. The CFD simulations were performed using ANSYS Fluent under identical operating conditions. The main findings of the study are summarized as follows:

- Modified flat-plate absorbers demonstrated a significant improvement in convective heat transfer and outlet air temperature compared to the plain configuration.
- The plain flat-plate absorber exhibited the lowest pressure drop but delivered the poorest thermal performance.
- The finned flat-plate absorber provided a favorable balance between thermal enhancement and hydraulic penalty, making it suitable for low- to medium-temperature solar drying applications.
- The V-grooved absorber achieved the highest outlet air temperature due to strong flow recirculation and turbulence; however, it also resulted in the highest pressure drop.
- Selection of the absorber configuration should therefore consider the trade-off between increased flow resistance and enhanced heat transfer in order to optimize energy consumption and drying efficiency.

Overall, the conclusions provide practical guidance for the design and selection of flat-plate air heaters used in solar drying systems. Future work should include experimental validation to confirm the CFD predictions. Additional studies may also investigate system performance under variable solar radiation and different boundary conditions representative of specific regions. Furthermore, optimization of groove geometry and fin arrangements could be explored to further enhance the thermal–hydraulic performance of modified flat-plate absorbers.

References

[1] Abd, H. M., Alomar, O. R., & Salih, M. M. M. (2022). Improving the performance of solar air heater using a new model of V-corrugated absorber plate having perforations jets. *International Journal of Energy Research*, *46*(6), 8130-8144.

[2] Baidhe, E., Clementson, C. L., Senyah, J., & Hammed, A. (2024). Appraisal of post-harvest drying and storage operations in Africa: perspectives on enhancing grain quality. *AgriEngineering*, *6*(3), 3030- 3057.

[3] Deepak, C. N., & Behura, A. K. (2023). Critical review on various solar drying technologies: Direct and indirect solar dryer systems. *Applied Solar Energy*, *59*(5), 672-726.

[4] Desisa, D. G., & Shekata, G. D. (2020). Performance Analysis of Flat-Plate and V-groove Solar Air Heater Through CFD Simulation. *International Journal of Renewable Energy Development*, *9*(3).

[5] Devan, P. K., Bibin, C., Shabrin, I. A., Gokulnath, R., & Karthick, D. (2020). Solar drying of fruits–A comprehensive review. *Materials Today: Proceedings*, *33*, 253-260.

[6] El-Sebaey, M. S., Mousavi, S. M., Sathyamurthy, R., Panchal, H., & Essa, F. A. (2024). A detailed review of various design and operating parameters affecting the thermal performance augmentation of flat- plate solar collectors. *International Journal of Ambient Energy*, *45*(1), 2351100.

[7] Kansara, R., Pathak, M., & Patel, V. K. (2021). Performance assessment of flat-plate solar collector with internal fins and porous media through an integrated approach of CFD and experimentation. *International Journal of Thermal Sciences*, *165*, 106932.

[8] Kalair, A. R., Seyedmahmoudian, M., Saleem, M. S., Abas, N., Rauf, S., & Stojcevski, A. (2022). A comparative thermal performance assessment of various solar collectors for domestic water heating. *International Journal of Photoenergy*, *2022*(1), 9536772.

[9] Oliveira, M., & Dutra, J. C. C. (2023). The impact of the V-corrugation on the thermal efficiency of a solar collector. *Solar Energy*, *255*, 460-473.

[10] Vengadesan, E., & Senthil, R. (2020). A review on recent developments in thermal performance enhancement methods of flat plate solar air collector. *Renewable and Sustainable Energy Reviews*, *134*, 110315.

Abrasive damage of mechanical face seals in Martian regolith

Benjámin ALEXE[1], Zoltán KÁROLY[2], Gábor KALÁCSKA[3]

[1] Institute of Technology, MATE, Gödöllő,
[2] Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, HUN-REN Research Network, Budapest
[3]Department of Materials Science and Engineering Processes, Institute of Technology, MATE, Gödöllő

Abstract:

During space missions, components and machinery must remain reliable and resilient even against the harsh environment of Mars, such as the abrasion caused by the aggressive Martian regolith. To achieve a reliable mechanical structure, protection of moving parts is a key factor. A highly effective solution for shielding against Martian regolith is sealing. Mechanical face seals represent a promising method with significant potential applications in the space industry. Due to the extreme temperature fluctuations in space, the application of solid lubricants is highly recommended. Among mechanical face seal materials, graphite is good solid lubricant and silicon carbide is very hard and resilient against abrasive regolith. Based on 24-hour tribological tests conducted according to Category IV of DIN 50322 standard, alongside morphological analysis via scanning electron microscopy (SEM), this mechanical seal face material combination performed as a practical solution in application at dry and abrasive environments. The performance is attributed to the solid lubrication and tribofilm formation of graphite, combined with the strength of silicon carbide.

Keywords:

mechanical face seals, three-body abrasive wear, solid lubricant, Martian regolith, space tribology

1. Introduction

Over the past decades, the exploration of Mars has been made possible by numerous successful vehicles. In the future, the surface exploration and research of Mars will potentially receive greater emphasis. Consequently, the effects of the surface environments of the celestial body on machinery used in space technology must be thoroughly investigated. Vehicles and structures operating on these celestial bodies must withstand highly unique stresses that are not experienced on Earth. On Earth, these stresses can be best simulated under

laboratory conditions, which allow for the adjustment of various parameters and environmental conditions. During these tests, as well as in already deployed vehicles and structures, unique engineering solutions are frequently applied to ensure adequate resilience and structural strength, or to maintain a low mass.

To achieve a reliable mechanical structure, the protection of moving parts is a key factor. Preventing Martian regolith dust from entering moving parts is crucial, as this dust is a primary factor that compromises the lifespan and reliability of machinery, such as a rover during Mars missions. A highly effective solution for shielding against Martian dusty and abrasive soil is sealing. Mechanical face seals represent a promising method with significant potential applications in the space industry.

These seals are commonly found in pumps, agitators, compressors, mixers or in any other machinery where a near-zero leakage sealing between machine components is required. Mechanical face seals are consists of two – one rotating and one stationary – flat and very smooth machined rings, that are pressed together by a combination of a spring force and the media's pressure, thereby creating a near-zero leakage sealing. The quality of the surface finish depends on the base material, but in general the average surface roughness is typically $Ra < 0.05$ to 0.1 μm for harder materials (such as ceramics), whereas for porous materials (such as carbon graphite), this value is $Ra < 0.1$ to 0.2 μm. The flatness of the face seals is inspected using optical methods and measured in light bands (typically helium light bands). For most industrial mechanical face seals, an acceptable flatness error is 1 to 3 light bands, which corresponds to 0.3-0.9 μm when using a helium light source. Due to these highly finished and flat surfaces, the mechanical face seal can ensure the formation of a micron-thick fluid film between the two faces, despite the axial preload (spring force) and the internal pressure. This fluid film enables the formation and maintenance of hydrodynamic, mixed, boundary, or elastohydrodynamic (EHD) lubrication regimes, thereby preventing dry running, severe wear, and can extend the seal's lifespan.

Due to the extreme temperature fluctuations in space, the application of solid lubricants is highly recommended. Among mechanical face seal materials, graphite is a good solid lubricant, and silicon carbide exhibits high hardness and resilience against abrasive regolith. Consequently, this mechanical face seal material combination can be a promising sealing solution between mechanical components in extremely abrasive, dry environments in the space industry [1].

2. Materials and methods

2.1. Carbon graphite material's properties (non-impregnated)

Carbon graphite material is highly common among mechanical face seal materials due to its advantages such as self-lubrication because of the porous structure, high chemical resistance, and good sealing property. The porous

structure of the base material also represents a drawback, it exhibits poor resistance to abrasive and crystalline materials and is prone to deformation in higher-pressure applications. Table 1 illustrates the carbon graphite base material's main properties.

Table 1. Carbon graphite base material's main properties [2] [3].

Carbon graphite	
Melting Point [°C]	3600 (sublimation)
Young's Modulus [GPa]	20-23
Compressive Strength [MPa]	120-180
Transverse Rupture Strength [MPa]	60-90
Vickers Hardness [-]	80-90
Density [kg/dm^3]	1.8-2.0

Furthermore, the thermal conductivity of carbon graphite is low, which is unfavourable regarding the dissipation of frictional heat. Additionally, its thermal shock resistance is not outstanding either (~9500 to 10,500 W/m). Moreover, as indicated by the table, it is a fundamentally a soft base material with a maximum Vickers hardness of 100, which is attributed to the material's porosity. The hardness, as well as the compressive and transverse rupture strengths, can be enhanced through impregnation. In the case of impregnated carbon graphite seal face, resin or antinomy (Sb) impregnation is commonly applied .

2.2. Silicon carbide material's properties

Alongside carbon graphite, silicon carbide is one of the most common materials for mechanical face seals. The main advantage is the outstanding hardness and resilience against abrasive media. However, due to its extreme hardness, silicon carbide is inherently brittle, making it particularly vulnerable to impact loads and sensitive to significant flatness errors. The main properties of silicon carbide are shown on Table 2.

Table 2. Silicon carbide base material's main properties [2] [3] [4].

Carbon graphite	
Melting Point [°C]	2830
Young's Modulus [GPa]	240-480
Compressive Strength [MPa]	700-3500
Transverse Rupture Strength [MPa]	100-750
Vickers Hardness [-]	2500-3000
Density [kg/dm^3]	2.4-3.3

Additional advantages include its high thermal shock resistance (~59,000 W/m) and good thermal conductivity, which, for silicon carbide it is in the range of 150 to 200 W/mK. Overall, it is an exceptionally durable base material [2] [5].

2.3. Experimental method

To investigate the SiC/Gr mechanical seal face combination in Martian regolith, an experimental test bench has been used, which is illustrated in Figure 1.

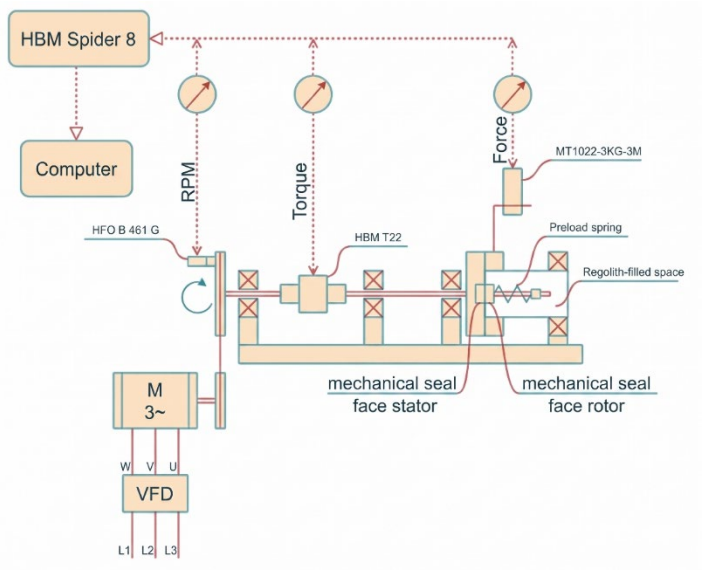


Figure 1. Schematic illustration of the experimental test bench.

During the experiments, dry (unlubricated) operating conditions were established to investigate whether the abrasive Martian regolith, upon entering the system, penetrates between the mechanical seal faces. Furthermore, if such ingress occurs, the study aimed to evaluate its effects on the seal surfaces from mechanical, morphological, and tribological perspectives.

The schematic diagram above illustrates the configuration of the test bench system, as well as the arrangement of its main components and sensors. The electric motor depicted in Figure 1 is controlled by a variable frequency drive (VFD) to regulate its rotational speed at 200rpm (+/-1%) drives a shaft, providing the torque input for the system. The three-phase electric motor transmits its torque and power to a shaft via a ribbed belt drive. The rotational speed of this driven shaft is measured at the pulleys using a B 461 G Hall-effect sensor manufactured by HFO.

The total torque is measured using a rotary torque transducer (HBM T22), which captures the cumulative frictional resistance of all moving elements. The frictional torque generated by the mechanical seal faces, located on the right side, represents a measurable torque loss within the system. This torque is transmitted through a bearing-mounted housing and a lever arm to a load cell (MT1022) with a 3kg maximal measurement range. The reaction torque lever arm pushes the load cell, which measures the reaction force; by knowing the length of the lever arm, the reaction torque originating from the friction of the mechanical seals can be calculated, allowing it to be isolated from the total system torque and other inherent torque losses. The lever arm is connected to the regolith-filled chamber which is shown in Figure 2.

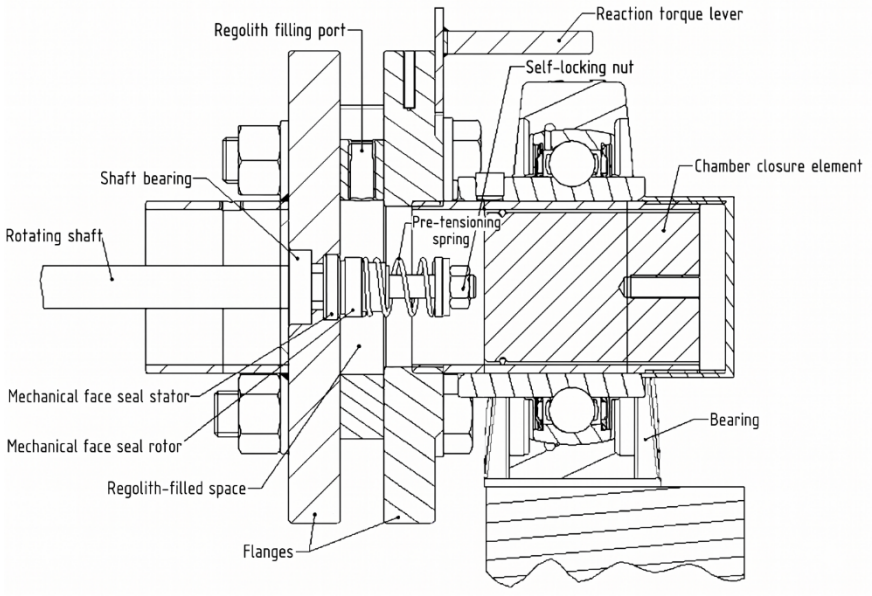


Figure 2. Test bench regolith-filled chamber cross section.

Figure 2 illustrates the cross-sectional view of the regolith-filled chamber. The figure shows the shaft, driven by the three-phase electric motor. The regolith-filled space is situated immediately after its bearing assembly. Within this section, the stationary and rotating components of the mechanical face seal are positioned between two flanges, and the driven shaft terminates at this point. The stationary ring of the mechanical seal is housed within the flange, which is supported by bearings on its opposite side. Consequently, during operation – as the through-shaft rotates – the friction between the two mating faces generates a corresponding frictional torque capable of rotating the flange.

The rotating component of the mechanical seal is mounted on the rotating shaft, and its sealing face is pressed against the stationary face by a spring. This spring allows for the application of various preload forces by adjusting its compression. The axial preload exerted by the spring remains constant throughout operation, as the rotating assembly and the spring are secured against axial displacement along the shaft by a self-locking nut located at the threaded shaft end.

During the measurements, rotational speed, torque, and force values were recorded using a HBM Spider 8 data acquisition (DAQ) module connected to a computer. The computer logged, stored, and processed the measured data according to the predefined sampling duration and cycle count. The experiment of each mechanical seal pair was conducted over an exact 24-hour period; therefore 48 sampling cycles were executed. Consequently, a 5-second sampling event was initiated every 1795 seconds (approximately every 30 minutes). The sampling frequency was set to 100 Hz, meaning that 100 data points were acquired per second per instrument, which resulted in a total of 500 data points per 5-second sampling cycle and 24,000 data points per 24 hours.

The testing of the SiC/Gr mechanical seal face in Martian regolith was conducted using two different spring preload settings. In the first experiment, the spring compression was set to theoretical zero, simulating a mechanical failure of the spring. In the second experiment, the preload spring was compressed by 8.6 mm, which corresponds to the nominal compression specified by the manufacturer of the mechanical seal face. Prior to these tests, the spring characteristics were determined in a separate study [6] with the result of approximately 2 N/mm. Based on the spring characteristic, the spring preload force can be determined for both configurations. At the theoretical zero compression, the axial force was generated by the surface friction of the mechanical seal faces, amounting to approximately 10.6 N. In contrast, at the 8.6 mm nominal compression, the applied preload force was 17 N.

2.4. Martian regolith

In this study, MGS-1 (Mars Global Simulant) was used in its unfractionated (bulk) form for both spring preload configurations. MGS-1 is a Martian regolith simulant manufactured by Exolith Lab, based on soil analyses conducted by the

Curiosity rover on Mars. The particle size distribution of the simulant is summarized in Table 3 [7] [8] [9].

Table 3. MGS-1 particle size distribution [7] [9].

Particle diameter	Finer particle diameter
1 mm	100 %
250 µm	92.6 %
125 µm	78.0 %
75 µm	59.6 %
45 µm	41.9 %
10 µm	14.3 %

The MGS-1 regolith is characterized by a basaltic mineral composition. The specific mineralogical components, expressed as weight percentages (wt%), are detailed in Table 4.

Table 4. Mineralogical components of the MGS-1 regolith [7] [9].

Mineral component	Content (wt%)
Anorthosite	27.1 %
Basaltic volcanic glass	22.9 %
Pyroxene	20.3 %
Olivine	13.7 %
Magnesium sulfate	4.0 %
Ferrihydrite	3.5 %
Hydrated silica	3.0 %
Magnetite	1.9 %
Anhidrite	1.7 %
Iron carbonate	1.4 %
Hematite	0.5 %
Összesen:	100 %

3. Results and discussion

As previous tests have [6] demonstrated, wear on the SiC mechanical seal faces was found to be negligible from a morphological perspective due to the inherent hardness of the base material. Consequently, the primary focus in this study is directed toward the graphite mechanical seal faces.

3.1. SEM morphological analysis of Gr mechanical seal face in MGS-1 regolith under theoretical zero preload

Scanning electron microscope (SEM) micrographs were acquired from the mechanical seal faces. These images clearly reveal morphological surface changes characteristic of three-body abrasion, alongside features indicative of plastic deformation, as illustrated in Figure 4.

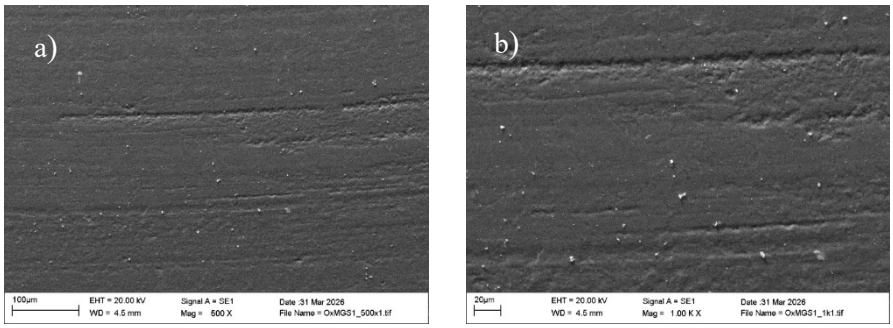


Figure 4. Acquired SEM images of graphite mechanical seal face tested in MGS-1 regolith under theoretical zero preload; a) 500x magnification, b) 1000x magnification

The micrographs reveals that abrasive wear is the dominant mechanism across the entire contact zone, evidenced by parallel ploughing grooves oriented in the sliding direction. The presence of these ploughing marks indicates that, alongside brittle morphological changes, mild plastic deformation also occurred on the surface.

Furthermore, as shown in Figure 4a, even at 500x magnification, the surface in the upper region exhibits the earliest signs of a flaking or lamellar texture. This is the morphological consequence of the incipient stages of a delamination wear mechanism, as described by Suh's theory of wear. During this process - driven by local Hertzian stress peaks and repetitive micro-level deformations (plastic deformation during groove formation, followed by flattening upon subsequent loading) - microcracks initiate. In this case, these micro-level deformations are more intense due to the inherently porous and layered structure of graphite. Based on the surface pattern of the rotor following further testing

with MGS-1 regolith, it can be concluded that the 24-hour test duration was insufficient for a fully distinct, intensive delamination to develop.

Moreover, the plastic deformation is even more pronounced in Figure 4b in the lower region. Consequently, it is evident that the self-lubricating properties of the graphite base material actively contributed to the contact mechanism during three-body abrasion, facilitating the formation of graphite tribofilm. This 1000x magnification micrograph provides a clearer visualization of the ploughing groove located in the upper region of the image. It is clearly evident here that, instead of classical micro-cutting, the embedded MGS-1 regolith particle induced plastic deformation by ploughing through the surface, consistent with the mechanism described earlier.

3.2. Evaluation of the acquired measurement data under theoretical zero preload

Figure 5 shows a graph about the mean reaction torque as a function of standard deviation and range.

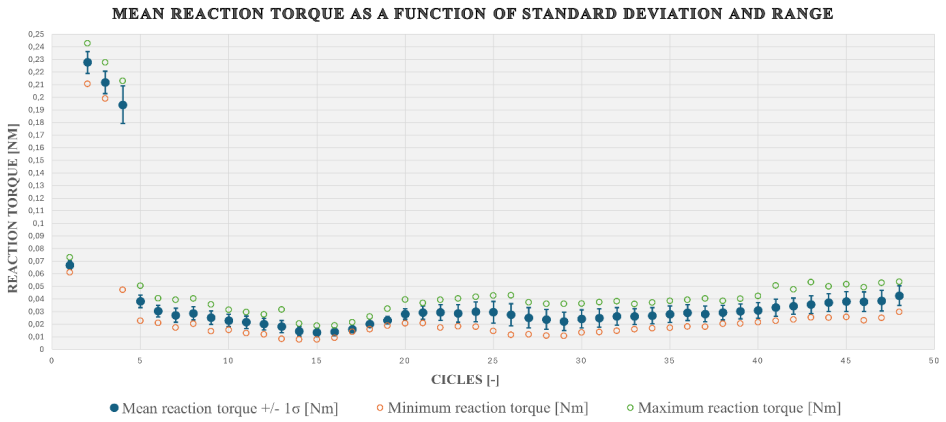


Figure 5. Mean reaction torque as a function of standard deviation and range (theoretic. zero preload)

Up to cycle 4 (hour 2), the mean reaction torque increases drastically, peaking at 0.22–0.23 Nm during cycle 2. This surge is attributed to the initial penetration and entrapment of MGS-1 particles between the seal faces, causing severe three-body abrasion via micro-cutting and abrasive ploughing on the graphite surface. By cycle 5 (hour 2.5), the mean torque sharply decreases to approximately 0.04 Nm. This drop, accompanied by a continuous decline in the torque range and standard deviation up to cycle 15 (hour 7.5), indicates the stabilization of plastic deformation and the successful formation of a graphite tribofilm. The concurrent fragmentation and blunting of the abrasive particles further reduce friction,

shifting the dominant wear mechanism from intensive micro-cutting to shear-induced plastic deformation. From cycle 15 onwards, a moderate but distinct increase in the mean torque is observed. This resurgence is driven by the partial detachment of plastically deformed surface layers and the reactivation of abrasive mechanisms. The absence of spring preload allows coarser regolith particles to more easily wedge the seal faces apart axially, permitting the ingress of new, larger MGS-1 particles into the contact zone. This re-intensification of micro-cutting and ploughing is evidenced by the progressive divergence of the maximum and minimum torque values (a widening range) from cycle 20 (hour 10) through cycle 30 (hour 15). Despite this, the overall torque values remain relatively moderate, suggesting that the entrapped particles also undergo rolling or rolling-sliding motion rather than pure sliding abrasion.

Figure 6 details the change of the reaction torque range values and illustrates the trend of the values with a trendline. In the graph, the initial torque range values exhibit drastic fluctuations, reflecting the instability of the system. By cycle 5 (hour 2.5) - coinciding with the sharp decrease in mean torque previously observed in Figure 5 - the system begins to stabilize, a behaviour also captured by the trendline. This progression is consistent with the previous diagram, as this stable phase persists for approximately the same duration, up to cycle 17 (hour 8.5).

Subsequently, as shown in Figure 6, the torque range begins to increase drastically up to cycle 26 (hour 13). This widening is driven by the re-emergence of the micro-cutting mechanism, which induces prominent extreme values, thereby increasing the overall range. From this cycle onward, a steady-state condition develops, resulting in no significant further fluctuations in the range. Approaching the final cycle, it is hypothesized that delamination initiates. Based on the increasing range values at this stage, it can be extrapolated that had the measurement period been extended, an upward trend would have followed, indicating renewed system instability.

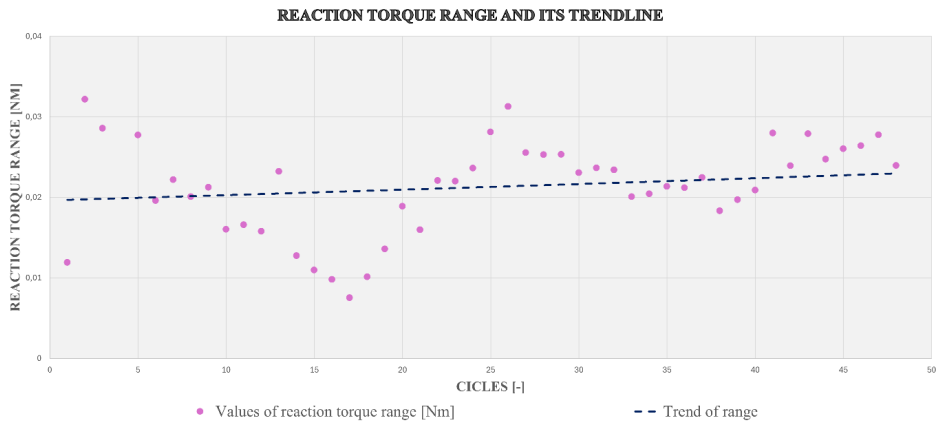


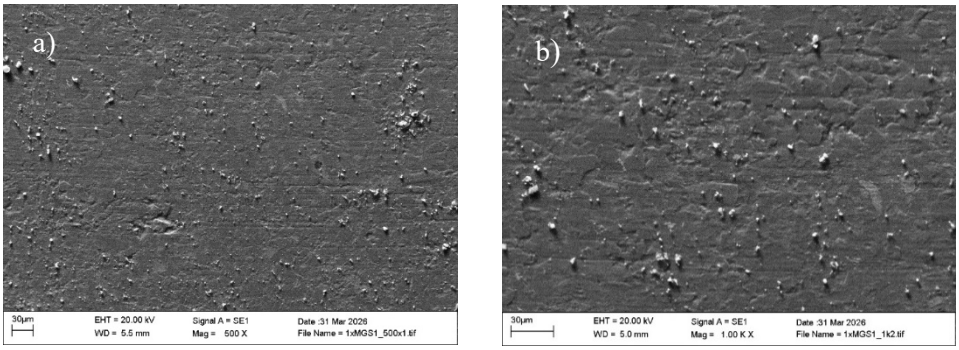
Figure 6. Reaction torque range and its trendline (theoretically zero preload)

3.3. SEM morphological analysis of Gr mechanical seal face in MGS-1 regolith under nominal preload

Scanning electron microscope (SEM) micrographs were acquired from the mechanical seal faces. These images clearly reveal morphological surface changes characteristic of three-body abrasion, alongside features indicative of plastic deformation, as illustrated in Figure 7.

Furthermore, Figure 7a clearly reveals that abrasive wear was the dominant mechanism across the entire contact zone, as evidenced by parallel micro-scratches and microgrooves oriented in the sliding direction. Driven by this abrasive action, micro-ploughing marks and mild plastic grooving developed on the surface. Additionally, an irregularly shaped depression or material pull-out can be observed in the lower-left central region in Figure 7a. Subjected to localized overloading (Hertzian stress peaks) induced by the regolith particles, the material in this area underwent surface fatigue, ultimately resulting in a fatigue-induced micro-pitting failure within the contact zone.

Figure 7. Acquired SEM images of graphite mechanical seal face tested in MGS-1 regolith under nominal preload; a) 500x magnification, b) 1000x magnification



Plastic deformation is evident in the upper and upper-right regions of Figure 7b. Based on the visual characteristics of this deformation - namely, flattened edges and a smeared texture - material smearing, indicative of tribofilm formation, can be concluded. Due to the inherently porous and layered structure of graphite, the load exerted by the abrasive particles is sufficient to cause a significant increase in shear stress within the surface plane. Consequently, surface fragments generated through spalling, material pull-out, and plastic pile-up during ploughing detach from the base material. Driven by the repetitive rotary sliding motion during the test, these detached wear debris particles are subsequently smeared back onto the graphite rotor surface by either the MGS-1 regolith particles or the SiC counterface.

3.4. Evaluation of the acquired measurement data under nominal preload

Figure 8 shows a graph about the mean reaction torque as a function of standard deviation and range. The mean reaction torque exhibits a steady increase up to cycle 8 (hour 4), reaching a local maximum of 0.085–0.09 Nm with a relatively constant torque range. This initial phase corresponds to the run-in period, where raw surface asperities interact with the MGS-1 regolith, initiating three-body abrasion via micro-cutting and abrasive ploughing. Between cycles 8 and 10, the mean torque briefly declines before resuming an upward trend with increased fluctuations (widening range and standard deviation) up to cycle 17 (hour 8.5). This instability indicates a transitional wear regime combining micro-cutting with plastic deformation. An 'action-reaction' loop emerges: abrasive material detachment is countered by material smearing and tribofilm formation, which plastically masks the underlying micro-grooves. Beyond cycle 17, the mean torque resumes an upward trajectory until cycle 30 (hour 15), accompanied by a diverging maximum-minimum range. This reflects the intensifying dominance of abrasive micro-cutting and material pull-out, although concurrent plastic deformation moderates the overall torque increase. Between cycles 30 and 35, the mean torque temporarily decreases while exhibiting extreme peak values. This signals a shift back toward plastic deformation dominance, occasionally disrupted by coarse particles causing severe ploughing spikes. Finally, from cycle 36 to 45, torque values rise again due to fatigue-induced micro-pitting and intensified abrasion, before declining toward the end of the 24-hour test (cycle 48).

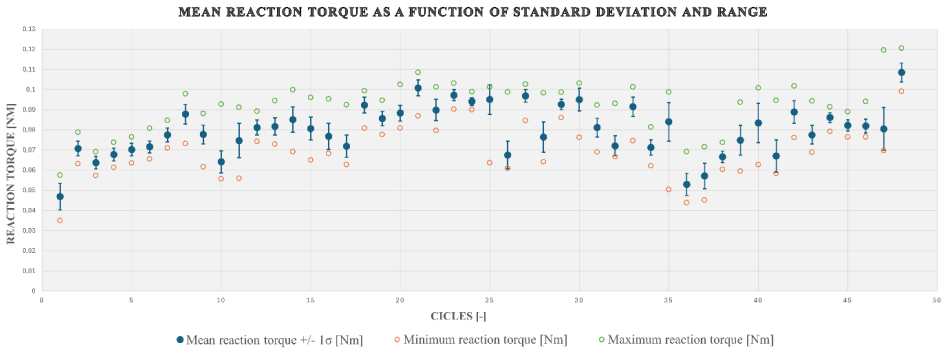


Figure 8. Mean reaction torque as a function of standard deviation and range (nominal preload)

Figure 9 details the change of the reaction torque range values and illustrates the trend of the values with a trendline. In the graph, the torque range exhibits minor spikes during cycles 1 and 2 (hours 0.5–1). This can be attributed to the initial ingress of MGS-1 regolith particles into the contact zone and the onset of localized micro-cutting and ploughing. From this point onward, the trendline reveals a continuous upward trajectory. Between cycles 5 and 7 (hours 2.5–3.5), the range stabilizes while the mean reaction torque increases, indicating steady

and continuous resistance from abrasive micro-cutting and ploughing. Subsequently, the range widens up to cycle 10 (hour 5). Correlated with the mean torque values, this widening signifies a transitional phase governed by a combined wear mechanism of plastic deformation and abrasive cutting. Thereafter, the torque range steadily declines until cycle 25 (hour 12.5). Given the concurrent rise in mean torque observed in Figure 9, this decline indicates the re-established dominance of the micro-cutting effect. Following this period, plastic deformation intensifies once again, leading to prominent extreme values and a subsequent expansion of the torque range.

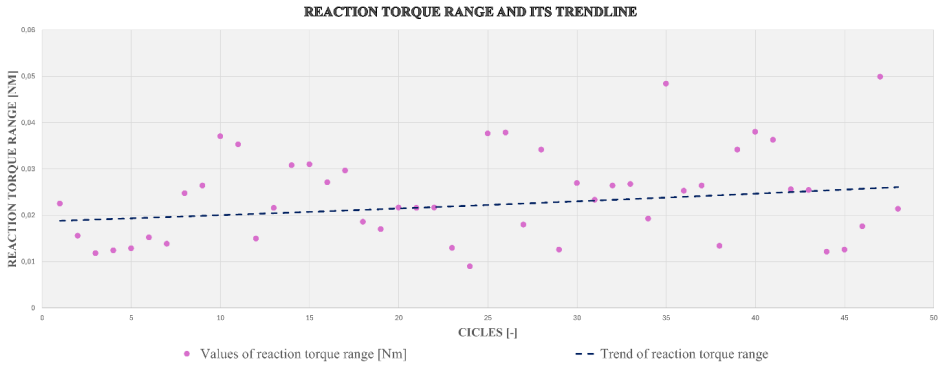


Figure 9. Reaction torque range and its trendline (theoretically zero preload)

4. Summary and conclusions

This study investigated the abrasive damage of mechanical face seals—specifically a Silicon Carbide (SiC) and Carbon Graphite (Gr) material combination—exposed to MGS-1 Martian regolith simulant under dry operating conditions. Based on the 24-hour tribological tests and subsequent morphological analyses, the following key conclusions can be drawn:

- SiC Durability: Wear on the Silicon Carbide mechanical seal faces was found to be negligible due to the inherent extreme hardness of the base material.
- Graphite Wear Mechanisms: The porous graphite seal faces experienced significant morphological changes dominated by three-body abrasion, micro-cutting, and plastic deformation. Furthermore, early signs of a flaking, lamellar texture indicated the incipient stages of a delamination wear mechanism.
- Tribofilm Formation: The self-lubricating properties of the graphite base material actively contributed to the contact mechanism, facilitating the

formation of a solid lubricating graphite tribofilm that helped stabilize friction during certain test phases.

- Performance Under “Zero” Preload: Simulating a spring failure (theoretical zero preload) allowed MGS-1 particles to easily penetrate the seal faces, causing an initial drastic surge in reaction torque due to severe micro-cutting. Although the system temporarily stabilized due to tribofilm formation, the absence of an axial spring force allowed coarser regolith particles to wedge the faces apart later in the test, reactivating abrasive micro-cutting and increasing system instability.
- Performance Under Nominal Preload: Under the manufacturer's nominal spring compression (17 N), the system exhibited a transitional wear regime characterized by an "action-reaction" loop. In this loop, continuous abrasive material detachment (spalling and ploughing) competed against the plastic smearing of wear debris and tribofilm formation. This constant abrasive interaction resulted in a steady overall upward trend in mean reaction torque throughout the test duration.

Ultimately, while the solid lubrication of graphite provides critical protection through tribofilm formation, maintaining adequate axial spring preload is essential to limit coarse particle ingress and ensure the reliable operation of moving parts in the highly abrasive and dry environment of Mars.

References

[1] G.A. Jones, "On the tribological behaviour of mechanical seal face materials in dry line contact: Part II. Bulk ceramics, diamond and diamond-like carbon films," Elsevier, Wear, Volume 256, Issues 3-4, p. 433-455, 2004, https://doi.org/10.1016/S0043-1648(03)00540-4.

[2] Sea-Land Srl and Eagleburgmann Germany GmbH & Co. KG, "kh.aquaenergyexpo.com," 11 2022. [Online]. Available: https://kh.aquaenergyexpo.com/wp-content/uploads/2022/11/Mechanical-Seal-Technical-Manual.pdf. [Accessed 11 01 2026].

[3] Mérnöki anyagok, Gödöllő: Szent István Egyetem, Gépészmérnöki Kar, 2018.

[4] Amechieneer, "amechieneer.com," 2022. [Online]. Available: https://amechieneer.com/properties-of-mechanical-seal-face-materials/. [Accessed 11 01 2026].

[5] S. Shankar, G. Praveenkumar and P. Krishna kumar, "Experimental study on frictional characteristics of tungsten carbide versus carbon as mechanical seals under dry and eco-friendly lubrications," Elsevier, International Journal of Refractory Metals and Hard Materials, Volume 54, p. 39-45, 2016, https://doi.org/10.1016/j.ijrmhm.2015.07.016.

[6] Kiss M., Kalácska G., Keresztes R. Zs. and Károly Z., "Tribological performance of mechanical face seals for Martian applications," Springer Nature, Scientific Reports p. 1-39, 2026, https://doi.org/10.1038/s41598-026-48495-2.

[7] Exolith Lab, MGS-1 Regolith Certificate of Analysis, USA FL: University of Central-Florida, Florida Space Institute, 2021.

[8] Kevin M. Cannon, Daniel T. Britt, Trent M. Smith, Ralph F. Fritsche and Daniel Batcheldor, "Mars global simulant MGS-1: A Rocknest-based open standard for basaltic martian regolith simulants," Elsevier, Icarus, Volume 317, p. 470-478, 2019, https://doi.org/10.1016/j.icarus.2018.08.019.

[9] E. Hütter, K. Seiferlin, P. Ehrenfreund, J. Garry, K. Gunderson, G. Kargl, A. Maturilli and J. P. Merrison, "Simulating Martian regolith in the laboratory," Elsevier, Planetary and Space Science, Volume 56, Issue 15, p. 2009-2025, 2008, https://doi.org/10.1016/j.pss.2008.09.017.

Vibration Sensors: principles, performance, and applications — a literature review

Tamim DEEB[1], László KÁTAI[2]

[1] Doctoral Program of Mechanical Engineering, MATE, Gödöllő
[2] Department of Machine Construction, Institute of Technology, MATE, Gödöllő

Abstract:

Vibration sensing underpins predictive maintenance, structural health monitoring, and an expanding share of safety-critical condition-monitoring systems. This review synthesises recent literature across the three dominant sensor families, optical methods (laser Doppler vibrometry and fibre Bragg grating sensors), micro-electromechanical and piezoelectric devices in capacitive and piezoresistive variants, and a heterogeneous group of eddy-current, electrodynamic, and emerging modalities that includes triboelectric nanogenerators, flexible piezoresistive films, and high-precision electrodynamic geophones, alongside the signal-processing and machine-learning stack that consumes their output. Three trends emerge. First, no sensor family dominates across applications. Each occupies a specific corner of the performance, environment, and cost space, and the correct comparator is always the target application rather than the sensor in isolation. Second, the boundary between signal processing and machine learning has effectively dissolved. Time-frequency representations are now produced primarily as 2D images for convolutional networks, and recent work shows convolution-attention hybrids and transfer learning across sensor types displacing pure CNN pipelines. Third, the remaining open problems are no longer transducer-physics problems but integration challenges: reporting standards, multi-modal fusion, edge inference, harsh-environment durability, and operator trust. The review covers performance metrics, deployment cost, comparative analysis, signal-processing methods, deep-learning approaches, and application domains from rotating machinery and structural health monitoring to manufacturing, agriculture, mining, and aerospace.

Keywords:

signal processing, vibration sensors, MEMS, piezoelectric, structural health monitoring (SHM), predictive maintenance, accelerometers

1. Introduction

Mechanical vibration is a routine signature of degradation in almost every machine that contains a rotating, sliding, or oscillating component. Bearings spall, gears chip, blades crack, and the resulting changes in the vibration signature appear long before the damage becomes catastrophic. Reliable vibration sensing, paired with a competent signal-processing pipeline, is the foundation of predictive maintenance, structural health monitoring (SHM), and an increasing share of safety-critical condition-monitoring systems. The economic case is straightforward: an unplanned shutdown of a cement-plant induction-draft fan, a wind-turbine gearbox, or a high-speed-rail traction motor costs orders of magnitude more than the instrumentation that would have caught the fault early (Khalil & Rostam, 2024; Cui et al., 2025).

The sensor landscape that supports this work has changed substantially in the past decade. Piezoelectric accelerometers, the historical workhorse of industrial vibration measurement, are no longer the obvious default. MEMS accelerometers have closed the performance gap for most predictive-maintenance use cases and dominate on size, cost, cabling, and (for digital variants) energy (Mohd Ghazali & Rahiman, 2021; Parisi et al., 2022). Optical methods, in particular laser Doppler vibrometry and fibre Bragg grating sensors, occupy specialised niches where non-contact measurement, distributed sensing, or immunity to electromagnetic interference matters more than absolute bandwidth (Bonopera, 2022; Shen, 2025). A long tail of specialist devices, electrodynamic geophones for sub-hertz seismic work, eddy-current arrays for in-situ structural inspection, triboelectric self-powered sensors, flexible piezoresistive films, fills the remaining corners of the design space (Kirchhoff et al., 2018; Song et al., 2021; Mehamud et al., 2022).

At the same time, the signal-processing and learning stack that consumes vibration data has shifted from hand-engineered features and shallow classifiers to time-frequency representations consumed by deep convolutional networks (Verstraete et al., 2017; Wei et al., 2025). The boundary between signal processing and machine learning has effectively dissolved: many recent papers produce 2D scalograms and spectrograms for the explicit purpose of CNN input (Lupea & Lupea, 2025; Gómez Muñoz et al., 2025). The implications run both ways. Sensor specifications can no longer be evaluated in isolation, because what matters is whether the device produces a signal that the downstream pipeline can decode. And the pipeline itself has become accessible enough that low-cost sensors deployed in agriculture, mining, and consumer-grade UAVs now feed credible diagnostic systems (Mystkowski et al., 2022; Alharbi et al., 2023; Al-Haddad et al., 2024).

2. Fundamentals and classification

Before turning to specific sensor families, it is worth fixing the working vocabulary that recurs across the literature: how vibration sensors are classified, what performance metrics they are evaluated against, and what counts as good practice in the data-acquisition pipeline. The thirteen review and methodology papers gathered for this purpose are unusually consistent on the taxonomy but disagree on what should count as the most important performance metric, which itself reflects the diversity of the application space.

2.1. Classification of vibration sensors

The reviews converge on a transduction-based taxonomy. Hassan et al. (2024) report that accelerometers dominate vibration acquisition, ranging from simple uniaxial models through to triaxial piezoelectric and MEMS variants, and that gearbox and bearing components are the primary data-gathering targets in the reviewed literature (Hassan et al., 2024). Huseynzade et al. (2024) extend the picture beyond accelerometers, noting that strain gauges remain valuable for detailed strain analysis, and that eddy-current sensors allow non-intrusive surface measurement on metal objects without contact (Huseynzade et al., 2024). Mohd Ghazali and Rahiman (2021) place these primary modalities in historical context: piezoelectric accelerometers remain the workhorse, MEMS units are gaining ground because of lower cost, and non-contact sensors such as LDV remove mounting problems altogether (Mohd Ghazali & Rahiman, 2021). Goyal and Pabla (2016) add a useful distinction by separating SHM techniques into local methods (sensitive to defects on the scale of the signal wavelength) and global methods (sensitive to overall structural response).

A second wave of reviews draws attention to the rapidly expanding set of "non-classical" data sources. Sony et al. (2019) argue that smartphones, UAVs, and consumer cameras now form a credible alternative to dedicated accelerometers for civil-infrastructure monitoring, using digital image correlation and motion magnification to extract vibration signatures remotely (Sony et al., 2019). The cost advantage and the ability to deploy via crowdsourcing change the economics of SHM, although image-based methods remain limited to lower frequencies. Garcia et al. (2025) frame the same evolution from a data-pipeline perspective, identifying where vibration analysis has already passed the deep-learning threshold and where it remains stuck on data scarcity or concept drift (Garcia et al., 2025).

2.2. Performance metrics

Sensitivity, bandwidth, noise floor, and dynamic range are the four metrics that recur across every review, but the literature treats them as a tightly coupled set rather than independent figures of merit. Zhang et al. (2022) note that time-domain feature extraction (AR, ARMA, PCA) is efficient for parameter estimation but tends to lose spatial information, while frequency-domain

methods (FFT, FRF, FDD, MUSIC) handle noisy signals and closely spaced modes better (Zhang et al., 2022). Aiordăchioaie (2025) moves the conversation from sensor specifications to algorithm performance, evaluating seven distinct fault-detection methods on the CWRU bearing dataset and concluding that time-frequency techniques (specifically short-time Fourier and Choi-Williams distributions) deliver the best classification, with the additional constraint that data windows above 500 samples are required for time-frequency methods and at least 1000 samples for frequency-domain methods to reach classification accuracies above 90% (Aiordăchioaie, 2025). Arts and van den Broek (2022) make a related point on the algorithm side with their fast continuous wavelet transform (fCWT), which outperforms the discrete wavelet transform on resolution and the short-time Fourier transform on frequency-range adaptability while remaining fast enough for real-time use (Arts & van den Broek, 2022). The implication is that headline sensor specifications can only be interpreted in the context of the signal-processing pipeline they will feed.

2.3. Quality, reproducibility, and noise

A persistent concern in the methodology papers is uncertainty quantification at the sensor level. Nichani et al. (2025) report that physical noise factors (electromagnetic interference, cable vibration, cable length) shift accelerometer outputs measurably, and propose a factorial design across eight measures of amplitude evaluated over varying time-window lengths to establish reliability thresholds before deployment (Nichani et al., 2025). Bointon and Leach (2020) approach the same problem from the hardware side, building an FPGA-based monitoring system that handles six triaxial accelerometers at up to 12,800 samples per second with on-the-fly peak detection (Bointon & Leach, 2020). Their CNC-lathe validation captured spindle-frequency components and showed that an open, programmable acquisition layer can serve research workflows that proprietary black-box systems cannot.

The strongest critique of the field's reporting standards comes from Bagri et al. (2024), who audited the whole vibration-analysis pipeline for rotating-machinery diagnosis and found that only 15.41% of recent papers report their preprocessing methods and only 10.15% include experimental validation (Bagri et al., 2024). Time-scale decomposition methods such as empirical mode decomposition and variational mode decomposition are now standard for handling non-linear and multi-component signals (Bagri et al., 2024). The most commonly deployed classifiers are SVM, CNN, LSTM, KNN, and random forests, with Transformer models emerging for remaining-useful-life prediction. If the preprocessing audit is right, however, much of this work is hard to reproduce, a problem that has more to do with reporting culture than with sensors or algorithms themselves.

The picture that emerges from this set of reviews is consistent. The taxonomy of sensors is well established and stable. The metrics are well understood individually but coupled in practice. The signal-processing pipeline that follows

the sensor matters as much as the sensor itself. And the reliability of the published results depends on disclosure practices that are not yet routine. The remaining sections of this review work through the first three points by sensor family.

3. Optical vibration sensing

Optical methods are the strongest competitor to inertial sensors when an application calls for non-contact measurement, immunity to electromagnetic interference, or measurement at points where attaching an accelerometer is impractical. The papers gathered here split naturally into three sub-families: laser Doppler vibrometry (LDV), fibre Bragg grating (FBG) sensors, and interferometric or full-field optical methods. The three are usually deployed in different situations rather than competing for the same one, because their failure modes do not overlap.

3.1. Laser Doppler Vibrometry

LDV uses the Doppler shift of light scattered from a vibrating surface to measure velocity directly, without contact and without mass-loading the structure. The technique itself has been mature for decades, and most recent work targets the signal processing problem that arises when the instrument is in motion. Darwish (2020) and (2022) addressed this from the time-domain and frequency-domain sides respectively. The 2020 paper reduced mean errors by 34.5 dB with a time-domain correction, against 25.3 dB for the standard frequency-domain method, an order-of-magnitude improvement (Darwish, 2020). The 2022 paper extended the analysis with a modified frequency-domain technique that produced a 33.5 dB error reduction below 100 Hz on stationary signals; for transient signals the time-domain method remained the only viable option, retaining its eightfold advantage over the frequency-domain approach (Darwish, 2022). Both methods rely on a reference accelerometer and careful sample-rate synchronisation. The practical implication is that LDV from a moving platform such as a UAV is now feasible, provided the correction algorithm is matched to the signal type.

LDV is also being used for acoustic emission monitoring in materials testing. Shen (2025) captured acoustic emissions during tensile tests of CFRP composites with LDV, isolating fibre-fracture emissions above 300 kHz that conventional piezoelectric AE sensors failed to capture because of their narrower frequency response. The absence of a coupling agent removes one source of signal distortion, at the cost of needing adequate surface reflectivity and tolerance to optical-path interference (Shen, 2025). Kalybek (2021) ran an experimental modal analysis comparing LDV, multi-camera systems, and consumer cameras against accelerometer references on building-scale structures. LDV reached a modal damping error of only 0.54% on mode 1, while image-

based systems matched the lowest modes well (RMS displacement error below 0.02 mm) but missed frequencies above roughly 20 Hz because of their higher noise floor (Kalybek, 2021). Jasiński (2025) carried the technique into musical acoustics, comparing LDV, magnetic pickups, and microphones on guitar strings. LDV gave the highest SNR at 49.1 dB and resolved every fine traveling-impulse feature, although it required a reflective target attached to the string that perturbed its vibratory energy (Jasiński, 2025).

3.2. Fibre Bragg Grating sensors

FBG sensors trade the high bandwidth and single-point focus of LDV for distributed measurement along a single optical fibre (Dalla Vedova et al., 2024). The basic mechanism is wavelength-encoded: each Bragg grating reflects a peak whose wavelength shifts with strain or temperature, and because the information is carried in wavelength rather than intensity, the readout is intrinsically immune to optical-power fluctuations (Dalla Vedova et al., 2024). Strain and temperature shift the same wavelength, so any FBG deployment has to consider how to disambiguate the two effects (Quattrocchi et al., 2021).

The seven FBG papers in this collection span low-frequency civil-infrastructure monitoring through to high-frequency aerospace work. Velázquez-Carreón (2026) reported a bridge-type FBG accelerometer with sensitivity up to 1730 pm/g in the 1–14 Hz band, transverse vibration error below 6%, a dynamic range of 58.74 dB, and an acceleration noise density of 122.62 μg/√Hz, all using 3D-printed PLA flexure hinges to keep costs low. The trade-off is the thermal and humidity sensitivity of the polymer, plus an unavoidable trade between bandwidth and maximum sensitivity (Velázquez-Carreón, 2026). At higher frequencies, Dalla Vedova (2024) used rapid prototyping to produce an FBG cantilever sensor with an almost flat response between 15 Hz and 200 Hz, with the first natural frequency of the support detected cleanly at 220 Hz. Signal amplification and distortion near that natural frequency are the main caveat, an issue that recurs in cantilever-based FBG designs in general (Dalla Vedova, 2024).

Lamberti (2016) and Bonopera (2022) addressed the interrogation problem rather than the sensor itself. Lamberti's VCSEL-based interrogator reached a dynamic precision of about 3 nε/√Hz at sample rates up to 50 kHz with 16 pm wavelength resolution, and estimated resonance frequencies within 0.2% of accelerometer references, all at roughly $4k. That is well below the cost of a tunable-laser interrogator, although optical power is restricted to below 1 mW and the scanning bandwidth is limited to about 9 nm (Lamberti, 2016). Bonopera (2022) reviewed FBG displacement-sensing variants and reported wedge-shaped slider configurations with sensitivities up to 123 pm/mm over 0–20 mm, large-range designs with linear correlation $R^2 = 0.9985$ over 0–200 mm, and intensity-based hole-core fibres with 8.16×10^{-6} mm resolution. The recurring weakness is strain-temperature cross-sensitivity in standard cantilevers and the chirp phenomenon in equal-section reflection spectra (Bonopera, 2022).

Algorithmic improvements have continued to compound. A fast phase-correlation (FPC) method with PolyMax estimation outperformed conventional maximum-detection for modal-frequency identification in four out of five experimental tests (Lamberti et al., 2015), although damping-ratio identification accuracy remained notably worse than frequency accuracy. de la Torre et al. (2021) demonstrated the durability of FBGs in underwater settings on hydraulic-turbine-related structures, achieving roughly 1 με strain accuracy and resolving subtle resonance shifts such as 6.3 Hz versus 7–8 Hz. The main field-level concerns the authors note are fibre fragility during installation and the need for specialised personnel (de la Torre et al., 2021).

3.3. Interferometric and full-field optical methods

The reviews by Gangopadhyay et al. (2004) and Rodríguez García et al. (2010) frame the landscape that more specialised methods sit in. Both authors observe that intensity-based fibre sensors are simple but vulnerable to optical-power fluctuations. Rodríguez García et al. (2010) add that FBG interrogation stages can resolve oscillations up to 400 kHz and that Fabry–Pérot interferometer (FPI) precision is limited by mirror-fabrication imperfections and alignment errors. Gangopadhyay et al. (2004) report an FPI displacement resolution better than 0.033 μm over a 20 μm range and reliable operation up to 2 kHz, with a 3.74 μm range under dynamic test, and note that interferometric methods generally face directional phase ambiguity that has to be resolved through additional processing.

4. Mems, Piezoelectric, Capacitive, and Piezoresistive sensors

This is the broadest family in the review. The fifteen papers cover four transduction principles and span price points from sub-dollar accelerometers in DIY IoT modules to monolithic triaxial devices designed for ballistic shock measurement. What ties them together is microfabrication, not a shared sensing mechanism. The discussion below is therefore organised by transduction principle, because the trade-offs within each are sharper than the trade-offs across families.

4.1. MEMS versus piezoelectric: where each wins

Five papers in this collection ran head-to-head comparisons between MEMS accelerometers and piezoelectric (IEPE) sensors, and they converge on a consistent picture. Symes et al. (2025) showed that MEMS units could measure the very low damping of aluminium accurately, where IEPE sensors introduced spurious damping errors attributable to their heavy cables. The trade-off the authors reported is that MEMS had a higher inherent background noise than commercial IEPE devices (Symes et al., 2025). Miranda et al. (2018) compared an ADXL203 against a piezoelectric reference on a shaker and saw matched

RMS tracking up to 2.0 g_RMS across 1 to 2500 Hz, with crest factor and kurtosis statistics that aligned closely with the piezo unit (Miranda et al., 2018). Encapsulation was the main practical constraint: bare PCBs do not survive machinery environments without custom protection (Miranda et al., 2018).

Ompusunggu et al. (2021) pushed this further with the ADXL1001 and ADXL1002 series, packaged in custom PCBs for industrial monitoring. The devices delivered ±3 dB bandwidths up to 10 kHz and dynamic ranges of 100 g and 50 g respectively, and successfully detected bearing faults in induction motors and blade failures in expander machines (Ompusunggu et al., 2021). Noise density remained higher than that of high-end piezo units, and custom epoxy potting was needed to survive harsh industrial settings (Ompusunggu et al., 2021). Parisi et al. (2022) examined digital and analog MEMS against piezo in the context of structural health monitoring, finding that digital MEMS in a low-power mode identified modal frequencies with a maximum error of only 1.6%, while consuming 40–80% less energy than full-power operation depending on the transmission protocol (Parisi et al., 2022). Their measured noise floor was 23.98 to 436.2 μg/√Hz across the MEMS devices tested, against 10.28 μg/√Hz for the piezo reference (Parisi et al., 2022). The cost of the energy savings was a sampling rate low enough to produce peak-amplitude estimation errors of up to 25%.

At the most cost-sensitive end of the market, Rodrigues et al. (2021) compared ADXL345 and MPU6050 modules under random engine vibration. The MPU6050 produced offset shifts under 0.03 g and resolved spectral components matching engine RPM, while the ADXL345 suffered vibration-rectification errors with offset shifts up to about 0.1 g and could not cleanly resolve the same spectrum (Rodrigues et al., 2021). Ultra-cheap modules can be viable for monitoring in a DIY IoT setting, but only if the device is chosen for its rectification behaviour and not just for its catalogue sensitivity.

4.2. Capacitive MEMS: pushing the noise floor

Three capacitive-MEMS papers attack the sensitivity-versus-noise trade-off in different ways. Xiong et al. (2025) used an anti-spring mechanism to soften the effective stiffness without resorting to bulky bias structures: their 4.2 × 4.9 mm sensor reached 51.1 mV/g sensitivity (a 10.4% gain), a noise floor of 21.3 μg/√Hz, bias instability of 0.24 mg, and nonlinearity of only 0.99% (Xiong et al., 2025). Fabrication tolerances are the limiting practical factor, since sidewall verticality and a stable DC bias both affect whether the anti-spring operating point lands where the design assumes (Xiong et al., 2025).

Silva et al. (2017) modelled the analytical side of the same problem, mapping how squeeze-film damping and thermoelasticity shape dynamic sensitivity across 0–1000 Hz and 228–398 K. Varying the packaging gas (air, argon, helium) and pressure between 20 Pa and 200 Pa moved damping and response significantly, which gives designers a way to optimise vacuum packaging before silicon is committed (Silva et al., 2017). The same paper includes an

experimental validation section with a developed electronic pick-off circuit and measurements that confirm the model predictions (Silva et al., 2017).

Li et al. (2024) took a different angle by encoding acceleration in time intervals of a resonating mass rather than in voltage or current. Sensitivity is tunable from −68.91 μs/g to −124.96 μs/g by adjusting vibration amplitude, with 1σ noise floors between 6.23 mg and 10.21 mg (Li et al., 2024). The approach removes some sources of ADC noise but is bottlenecked by the clock resolution of the timer, and the current prototype still relies on a conventional capacitance-to-voltage front end that adds hybrid noise (Li et al., 2024).

A related class is the micro-opto-electro-mechanical (MOEMS) accelerometer. Taghavi et al. (2022) closed an electrostatic feedback loop around a Fabry-Pérot optical cavity, reaching ±5 g range, 1.6 V/g sensitivity, 25.8 μg bias stability, 500 Hz bandwidth, and a 1403 Hz mechanical resonance (Taghavi et al., 2022). The closed-loop topology removes the usual trade-off between optical linear range and sensitivity, and the optical interrogation is immune to EMI, at the cost of an external laser and fibre-optic setup (Taghavi et al., 2022).

4.3. Piezoelectric MEMS: PZT versus PVDF

Three papers explore piezoelectric MEMS by varying the active material. Gong et al. (2023) used aerosol deposition to build thick PZT sensing layers, achieving 22.74 pC/g charge sensitivity, 4.96 mV/g voltage sensitivity, a 5% flat band from 10 to 200 Hz, and a noise-equivalent acceleration of 5.6 μg/√Hz, matching the SNR of commercial piezo references without the vacuum packaging usually needed for a high Q-factor (Gong et al., 2023). The unavoidable limitations are the thermal-electrical noise inherent in piezoelectric capacitance and the need for a charge-amplifier conditioning circuit (Gong et al., 2023).

Ge et al. (2023) and (2024) pushed PVDF as a lead-free alternative. The 2023 paper combined PVDF films with rapid 3D printing and laser micromachining to yield 21.82 pC/g charge sensitivity, 126.32 mV/g open-circuit voltage sensitivity, a 5% flat band up to 58.5 Hz, and noise density of 6.02 μg/√Hz, targeting fully organic microsystems for wearable bio-monitoring while still relying on copper-layer electrode interconnects and silver-paste connections in the actual prototype (Ge et al., 2023). PVDF's intrinsically weaker piezoelectric coefficient forced larger structural dimensions and capped resonance at 126.46 Hz (Ge et al., 2023). The 2024 follow-up extended performance with a multi-unit trapezoidal cantilever array: average sensitivity rose to 29.45 pC/g over ±10 g, the 5% flat band extended to 160 Hz with fundamental resonance at roughly 422 Hz, and in-band noise dropped to 1.4 μg/√Hz. The trade-off is footprint (15 × 15 mm) and manual lamination, which currently caps scalability (Ge et al., 2024).

4.4. Piezoresistive triaxial designs

Yu et al. (2023) and (2025) built triaxial piezoresistive accelerometers with a synchronised-deformation mechanism designed to deliver position-independent pure axial stress. The 2023 paper covered ±100 g with x/y/z sensitivities of 2.43, 2.44, and 2.41 mV/g/5V, natural frequencies of 11.4 kHz on the in-plane axes and 13.2 kHz on the vertical axis, and cross-axis sensitivity below 1.56% (Yu et al., 2023). The temperature-coefficient drift on the z-axis was 0.00118 FS/°C, which the authors note is comparable to commercial products and does not affect downstream applications, with packaging effects accounting for some residual shift in resonance against the simulated target (Yu et al., 2023). The 2025 paper extended the design to extreme shock measurement, reaching 200,000 g range with natural frequencies above 1.5 MHz and sensitivities of about 1.165 to 1.223 μV/g/5V at nonlinearity under 2.81% (Yu et al., 2025). This is the kind of dynamic range needed for blast-fuze and penetrating-weapon instrumentation, and the fabrication relies on bulk micromachining steps including DRIE, anodic bonding, and laser etching (Yu et al., 2025).

4.5. Multi-bandwidth integration

Tsai et al. (2021) made a different choice: rather than design a single sensor that spans a wide bandwidth, they packaged five commercial MEMS chips spanning three types (one ADXL327, one ADXL203, and three ADXL001 units) with bandwidths of 1.5 kHz, 3.2 kHz, and 5.6 kHz into a 22 × 22 × 13 mm module. The composite covers a frequency range that no single low-cost MEMS device handles, and the module was small enough to embed near CNC spindle bearing seats and robot-arm joints, where it detected chatter and rotational vibrations effectively (Tsai et al., 2021). The redundancy comes at the price of eighth-order Butterworth filtering to clean overlapping spectra, and the ADXL001's low native sensitivity (25 mV/g) forces a signal-amplification stage that adds stationary noise to its channel (Tsai et al., 2021).

5. Eddy-current, Electrodynamic, and emerging vibration sensors

This category collects vibration sensors that fall outside the two dominant lineages of optical and inertial MEMS devices. The five papers cover an electrodynamic geophone built for low-frequency seismic work, two eddy-current devices that double as non-destructive testing instruments, a triboelectric nanogenerator (TENG), and a flexible piezoresistive film. None of these is positioned as a general-purpose accelerometer. Each targets a specific corner of the design space: sub-nanometer noise floor, conformability, self-powered operation, or low-cost flexible fabrication.

5.1. Electrodynamic geophones

For sub-hertz seismic measurement, the electrodynamic geophone is still the standard instrument. Its coil-and-magnet harmonic oscillator produces an inductive readout that can be driven close to the thermal Johnson limit if the amplifier is matched carefully (Kirchhoff et al., 2018). For an L-4C geophone paired with an OPA188 preamplifier, Kirchhoff et al. (2018) reported a noise floor of about 10^{-11} m/√Hz at 1 Hz over a 0.01–100 Hz bandwidth, which is what active seismic-isolation systems for gravitational wave interferometers need. The constraints on this performance are tightly coupled. The floor is set jointly by coil Johnson noise and amplifier voltage noise, so optimising one in isolation buys nothing. The floor is also so low that it cannot be measured directly: characterisation requires huddle-testing several units against each other in a seismically isolated environment (Kirchhoff et al., 2018).

5.2. Eddy-current sensing

Two papers use eddy-current transduction in very different ways. Song et al. (2021) printed a flexible eddy-current array (FECA) on a flexible PCB, bonded it conformally to aircraft metal structures with specialised aviation sealant, and tracked fatigue cracks roughly 1 mm to 3.2 mm long. The reliability data is unusual for sensors at this stage of development: signal variation stayed below 1% after 10^7 vibration cycles at 1055 Hz under 0.8 g to 8 g excitation, and the array also survived salt-fog corrosion, fluid immersion, and UV exposure (Song et al., 2021). High-frequency vibration is normally a worry for eddy-current sensors because it modulates the lift-off distance to the substrate, but the aviation-sealant bonding scheme suppressed that modulation in the FECA's reported experiments, which is part of what gives the sensor its stable in-service output (Song et al., 2021).

Xu et al. (2020) went the other direction by miniaturising the transducer rather than scaling out the array. Their spin-Hall magnetoresistance (SMR) sensor is driven by two AC currents, a bias current through the SMR element and an excitation current that injects the eddy-current probe field, to pick up the magnetic signature of secondary eddy currents, with a low-field detectivity near 1 nT/√Hz at 1 Hz and a linear sensitivity of 1.10 mV/V/Oe over ±0.84 Oe. The AC-driven topology produces a DC output directly, so there is no lock-in amplifier or external demodulator (Xu et al., 2020). With a PCA and decision-tree classifier on top, the sensor reached 100% accuracy on crack-type classification in conductive samples, though subsurface defects attenuated the signal heavily and the raw waveforms across crack categories were too similar to separate without that classifier (Xu et al., 2020).

5.3. Triboelectric (TENG) self-powered sensors

A TENG sensor solves a different problem: it generates its own output through mechanical contact and slip and so needs no external power supply. Mehamud et al. (2022) reported a spring-assisted TENG combining contact-separation and in-

plane sliding modes, with frequency response from 0 to 1200 Hz and acceleration range up to 120 m/s². The device reached peak outputs of 200 V and 0.9 μA, levels that the authors describe as highly desirable for a stable self-powered sensor (Mehamud et al., 2022). The spring-assisted geometry is the main reason the bandwidth is that wide. Bandwidth has historically been the weak point of TENG vibration sensors (Mehamud et al., 2022). One remaining issue reported in the paper is that single-spring topologies are slightly non-linear and less stable at higher frequencies than two-spring variants (Mehamud et al., 2022).

5.4. Flexible piezoresistive films

Flexible piezoresistive sensors are becoming competitive for low-cost machinery vibration monitoring. Mou et al. (2025) paired a micro-grooved PDMS membrane with a bouncing stainless-steel ball, so that ball impacts modulate membrane resistance. The device covers 50 Hz to 400 Hz across a 1 g to 5 g acceleration range, with a linear voltage-acceleration response ($R^2 = 0.92$) and response and recovery times under 50 ms at 10 kPa (Mou et al., 2025). Its omnidirectional response captures vibrations arriving from any angle, and the simple microstructural design fabricated with a femtosecond-laser template keeps the device cheap to produce (Mou et al., 2025). Sensitivity drops at higher pressures (50–150 kPa), and voltage amplitude drops at higher frequencies except for a resonance spike near 200 Hz (Mou et al., 2025).

6. Comparative analysis

The previous three sections covered the major sensor families in depth, but treated each on its own terms. This section places them side by side along the dimensions that actually matter when choosing a sensor for a new deployment: sensitivity and noise floor, bandwidth, dynamic range, mounting and environment, cost and integration burden, and the constraints imposed by the downstream signal-processing and learning pipeline.

6.1. Performance envelopes

The four conventional performance metrics, sensitivity, bandwidth, noise floor, and dynamic range are tightly coupled, and improving one usually costs another. The numbers reported across this review make the trade-offs explicit. Electrodynamic geophones reach noise floors near 10^{-11} m/√Hz at 1 Hz over a 0.01–100 Hz bandwidth but cannot operate above roughly 100 Hz (Kirchhoff et al., 2018). Capacitive MEMS accelerometers using anti-spring softening reach 21.3 μg/√Hz on a 4.2 × 4.9 mm die with sensitivity of 51.1 mV/g (Xiong et al., 2025). Piezoelectric MEMS using aerosol-deposited PZT reach 5.6 μg/√Hz with 22.74 pC/g charge sensitivity (Gong et al., 2023), and polymeric PVDF cantilever arrays push this further to 1.4 μg/√Hz on a larger 15 × 15 mm

footprint (Ge et al., 2024). At the extreme end of dynamic range, piezoresistive triaxial designs measure shocks up to 200,000 g with natural frequencies above 1.5 MHz (Yu et al., 2025).

Optical sensors map onto this space differently. LDV is fundamentally non-contact and can resolve modal damping with 0.54% error on building-scale structures and acoustic emissions above 300 kHz on materials testbeds (Kalybek, 2021; Shen, 2025), but its bandwidth is bounded by Doppler signal quality rather than by mechanical resonance. FBG sensors trade bandwidth for distributed sensing along a single fibre, with current designs reaching sensitivities up to 1730 pm/g in the 1–14 Hz band (Velázquez-Carreón, 2026) and FPI variants delivering sub-micrometre displacement resolution (Gangopadhyay et al., 2004).

6.2. Mounting and environment

Heavy IEPE piezoelectric units introduce spurious damping through their cables, which is invisible on a clean lab shaker but appears as systematic error when measuring the low damping of lightweight materials (Symes et al., 2025). MEMS sensors with their lighter mass and shorter cables side-step that problem at the cost of a higher noise floor (Symes et al., 2025; Parisi et al., 2022). For applications where any mass loading would distort the measurement, LDV is the only practical option, although it imposes its own constraints on surface reflectivity (Shen, 2025).

In harsh environments, the ranking changes again. FBGs survive underwater deployment on hydraulic-turbine-related structures with roughly 1 με strain accuracy (de la Torre et al., 2021). Flexible eddy-current arrays survive salt-fog corrosion, fluid immersion, and ultraviolet exposure over 10^7 vibration cycles (Song et al., 2021). MEMS modules require custom epoxy potting and protective casing to survive harsh industrial environments, which adds non-trivial integration cost (Ompusunggu et al., 2021; Miranda et al., 2018). Triboelectric devices are self-powered but wear over long-term operation because of the contact-slip transduction mechanism.

6.3. Cost and integration burden

The cost gap between sensor classes is now large enough to shape application choices on its own. Ultra-low-cost MEMS modules (ADXL345, MPU6050) are available for a few dollars and can carry serviceable diagnostic information if the device is chosen for its rectification behaviour (Rodrigues et al., 2021). The VCSEL-based FBG interrogator developed by Lamberti (2016) reaches commercial-grade dynamic precision at roughly $4k, well below the cost of a tunable-laser interrogator. LDV systems and electrodynamic geophones remain at the high end of the cost distribution, and digital Fresnel holography systems remain in the research-equipment tier. The cost-versus-performance frontier is therefore not a single curve but a stepped function: the jumps from MEMS to high-end piezoelectric to LDV are large and well separated.

6.4. The pipeline matters as much as the sensor

A theme that emerged in Section 2 and recurs throughout the review is that headline sensor specifications cannot be interpreted in isolation. The signal-processing pipeline determines whether a noise floor measured in μg/√Hz translates into useful diagnostic information. Aiordăchioaie (2025) showed empirically that frequency-domain methods need windows of at least 1000 samples to reach 90% classification accuracy, while time-frequency methods need windows above 500 samples, regardless of which sensor produced the data. Parisi et al. (2022) reported that a MEMS device in low-power mode produced peak-amplitude errors of up to 25% because of reduced sampling rate, even though the underlying noise floor was acceptable. These results imply that "best sensor for the job" is a question that has to include the pipeline.

6.5. Decision framework

A pragmatic selection process that emerges from this literature has roughly four steps. First, characterise the application by frequency band, dynamic range, environmental severity, and budget. Second, eliminate sensor classes that fail any hard constraint (sub-hertz work eliminates almost everything except geophones and FBG; harsh environments eliminate bare MEMS; mass-sensitivity eliminates contact sensors). Third, choose the cheapest remaining device whose noise floor is consistent with the downstream classifier's signal-to-noise demands. Fourth, design the signal-processing and learning pipeline alongside the sensor selection rather than treating it as a downstream concern. The literature is consistent in showing that the boundaries between the sensor, the signal-processing stack, and the learning algorithm have become porous, and treating them as independent choices is an outdated habit.

7. Signal processing for vibration analysis

The transducer is rarely the rate-limiting step in a vibration-monitoring pipeline. What separates a useful measurement from a noisy time series is the signal-processing stack that sits on top of it. The papers in this section cover that stack from raw FFT-based spectral analysis through to time-frequency representations that double as image inputs for deep neural networks. Three observations emerge across the literature. Adaptive decomposition methods (EMD and its descendants) have become the default first pass for non-stationary signals. Time-frequency methods (STFT, CWT, Wigner-Ville variants) are now the workhorses for fault diagnosis. And the boundary between signal processing and machine learning has dissolved: many recent papers produce 2D images explicitly so a CNN can consume them.

7.1. Frequency-domain foundations

FFT and spectral averaging remain the entry point for most vibration pipelines. Dragos et al. (2024) used cross-spectral density phase to formalise "time-lag" discrepancies in SHM data and applied vector spectral averaging, which cancels noise components rather than merely reducing their variability as RMS averaging does before further analysis (Dragos et al., 2024). Han et al. (2021) combined FFT with PCA in a more conventional pipeline that reduces dimensionality before classification: the FFT carries the diagnostic information, and PCA discards low-variance components before the data reaches an SVM or CNN classifier (Han et al., 2021). At the simplest end of the spectrum, the practical workhorse is still Parseval-theorem-based energy extraction from FFT bins, which Khalil and Rostam (2024) used to isolate specific fault bands for semi-automated fault diagnostics in rotating machinery (Khalil & Rostam, 2024).

7.2. Adaptive decomposition: EMD and its descendants

The major weakness of standard FFT is its assumption of stationarity, which mechanical fault signals routinely violate. Empirical mode decomposition (EMD) addresses this by decomposing a signal into intrinsic mode functions (IMFs) without imposing a basis. Rai and Mohanty (2007) demonstrated the canonical use case by generating IMFs in the time domain, then applying FFT to each individually, which removed the need to estimate physical time intervals for characteristic defect frequencies (Rai & Mohanty, 2007). Peng et al. (2005) addressed several practical EMD problems — non-monocomponent initial IMFs, low-frequency artefacts, and the masking of low-energy components, by using wavelet packet transform (WPT) as a preprocessor and applying EMD to the resulting narrow bands (Peng et al., 2005).

Two further families have grown around the same problem. Empirical wavelet transform (EWT), used by Kedadouche et al. (2016), segments the Fourier spectrum to build an adaptive wavelet filter bank, which mitigates EMD's mode-mixing problem and EEMD's computational cost simultaneously. A kurtosis-based statistical index selects the set of relevant IMFs for envelope-spectrum extraction (Kedadouche et al., 2016). Zhang et al. (2022) reviewed the broader landscape (EEMD, CEEMDAN, VMD, EWT) and highlighted that VMD resolves EMD's lack of mathematical foundation by iterating toward optimal variational modes, while deconvolution methods such as MED and MCKD recover periodic fault pulses buried in heavy noise (Zhang et al., 2022). The most recent survey of this space by Zhou and Tang (2025) emphasises successive VMD (SVMD) and feature mode decomposition (FMD), which iteratively update FIR filters using correlated kurtosis, as the current frontier (Zhou & Tang, 2025).

7.3. Time-frequency representations

Where adaptive decomposition gives a one-dimensional time-series output per mode, time-frequency representations produce a 2D map that humans and neural networks can both read. Verstraete et al. (2017) ran STFT, wavelet transform, and HHT side by side for rolling-bearing diagnosis, with the explicit goal of using the resulting spectrograms as image inputs to a CNN (Verstraete et al., 2017). Wei et al. (2025) extended the comparison to include modified S-transform and smoothed pseudo Wigner-Ville distribution (SPWVD), noting that SPWVD's independent time-frequency smoothing windows are the main reason it suppresses the cross-term interference that limits standard WVD (Wei et al., 2025). Pacheco-Chérrez et al. (2022) showed that even without going to 2D, the discrete wavelet transform can produce useful low-dimensional features. On bearing faults in rotating machinery they evaluated three distinct feature pipelines: 14 multi-domain statistical parameters fed to an SVM, relative wavelet energy features from the DWT, and a Linear Discriminant Analysis pipeline, each evaluated separately rather than fused into a single feature set (Pacheco-Chérrez et al., 2022).

7.4. Hybrid and filtering approaches

Several papers compare classical filtering against frequency-domain decomposition. Al-Haddad et al. (2024) ran a comparative analysis of discrete Kalman filtering against multi-level DWT for multirotor-UAV fault diagnosis, with the DWT-based multi-resolution analysis recovering fault-related content that the Kalman stage missed (Al-Haddad et al., 2024). Garcia and Tcherniak (2019) introduced singular spectrum analysis as an alternative to PCA that accounts for intercorrelation between signal vectors, capturing rotational patterns in wind-turbine blades that PCA misses (Garcia & Tcherniak, 2019).

8. Machine learning and deep learning in vibration analysis

The machine-learning side of the literature divides along two axes: what representation of the vibration signal is fed to the model, and what model class consumes it. Within the last decade, the field has moved decisively toward 2D time-frequency images consumed by convolutional networks, with classical ML methods retreating into niches where feature engineering is feasible and computation is constrained. The reviews by Zhang et al. (2022), Tama et al. (2022), and an unattributed 2023 anomalous-sound study cover the broader trajectory; the empirical papers fill in what has worked.

8.1. Classical ML on engineered features

The classical pipeline extract a feature vector from the signal, then classify with SVM, random forest, or LDA, remains a strong baseline. Pacheco-Chérrez et al. (2022) achieved over 96% accuracy on rolling-bearing fault detection using

statistical features from time, frequency, and time-frequency domains combined with SVM, random forest, and LDA classifiers (Pacheco-Chérrez et al., 2022). Khalil and Rostam (2024) reported 93.24% accuracy on rotating-machinery fault detection in industrial cement plants using a stacking ensemble of an MLP and SVC trained on FFT-spectrum features such as peak amplitudes and band energies, outperforming a single-SVC baseline of 86% (Khalil & Rostam, 2024). Ayankoso et al. (2024) made the comparison explicit: on induction-motor bearing faults and misalignments, random forests and XGBoost trained on FFT and DWT features were competitive with deep CNNs trained on raw data (Ayankoso et al., 2024). The implication is not that classical methods are obsolete but that, given good engineered features, they can match or beat deep networks on smaller datasets.

8.2. Convolutional networks on time-frequency images

By far the most common deep-learning recipe in this collection is to convert a vibration signal into a 2D time-frequency image and feed it to a CNN. Verstraete et al. (2017) set the template by comparing STFT spectrograms, wavelet scalograms, and HHT plots as 96 × 96 image inputs to a deep CNN. On rolling-bearing data they reached 99.9% on MFPT and 99.5% on CWRU (Verstraete et al., 2017). Lupea and Lupea (2025) took the scalogram approach further with concatenated CWT grayscale images fed to a 2D CNN for helical gearbox defect detection, reaching 99.91% versus 99.63% for the same CNN on raw signals — a small margin, but a consistent one (Lupea & Lupea, 2025).

Wei et al. (2025) ran the broadest head-to-head comparison of representations, evaluating STFT, CWT, modified S-transform, SPWVD, and HHT as inputs to a lightweight CNN. Modified S-transform delivered 99.70%-99.87% on bearing datasets, while STFT spectrograms held competitive ground at 98.77%-99.36% on gear datasets (Wei et al., 2025). The takeaway is that, with a well-tuned CNN, the choice of time-frequency representation matters but not enormously; what matters more is whether the representation captures the diagnostic signature cleanly. Pavithra and Ramachandran (2025) reinforced this with a DCNN trained on STFT and HHT-EMD-IMF representations that retained 100% classification accuracy on bearing faults under additive Gaussian noise up to 10 dB, without dedicated noise layers (Pavithra & Ramachandran, 2025).

8.3. Transfer learning and pre-trained backbones

Transfer learning is now standard practice when training data is limited. Quiles-Cucarella et al. (2025) used SqueezeNet with CWT scalograms generated by amor and morse wavelets and reached 99.37% on rolling-element bearings, with amor producing the better features (Quiles-Cucarella et al., 2025). Gómez Muñoz et al. (2025) demonstrated a more interesting transfer: they trained GoogLeNet on CWT scalograms from full-field laser vibrometer measurements and successfully transferred the learned features to cheap accelerometer data on

metallic plates, reaching near-100% accuracy (Gómez Muñoz et al., 2025). This is a meaningful result because it inverts the usual cost structure: the expensive instrument is used to teach the model, after which the cheap instrument can be deployed at scale. Łuczak (2024) combined six CWT scalograms, one per axis of a 6-axis IMU into a single RGB image for the CWTx6-CNN method, with a separate STFTx6-CNN model trained for comparison, and added Grad-CAM and LIME interpretability tools to process each fault classification in under 50 ms (Łuczak, 2024). Chennana et al. (2025) evaluated VGGish, VGG16, and YamNet individually on log-Mel spectrograms, then fused the features of the top-performing CNN with a hand-crafted MBH-LPQ texture descriptor via weighted summation, reaching 98.95% on CWRU and 100% on the PU bearing dataset (Chennana et al., 2025).

8.4. 1D-CNN and raw-signal approaches

Not all deep-learning approaches need a 2D image. Kolar et al. (2020) treated raw 3-axis accelerometer signals as a wide 6400 × 1 × 3 "image" for a multi-channel deep CNN, dispensing with handcrafted preprocessing entirely for rotary-machinery fault diagnosis (Kolar et al., 2020). Zhou et al. (2025) make the case for combining 1D-CNN with adaptive signal-processing techniques such as EMD and wavelet transforms on the grounds that lightweight networks become more physically interpretable when handed inputs that already concentrate fault-related features (Zhou et al., 2025).

8.5. Transformers, GANs, and hybrid models

Recent papers point to the next phase of method development. Wang et al. (2025) introduced an LGAF-Swin Transformer trained on WAA-FMD-filtered CWT maps, reaching 100% on the KAIST open dataset and 98.13% on measured bearing data, competitive with the best CNN results and notable for integrating depthwise separable convolutions alongside attention to extract multi-scale features (Wang et al., 2025). Faysal et al. (2021) addressed the data-scarcity problem with a deep convolutional GAN trained to augment NEEEMD and CWT scalograms; the baseline CNN reached 98.0% on bearing fault classification and 96.31% on blade fault classification, with DCGAN augmentation improving validation and test robustness further (Faysal et al., 2021). Moysidis et al. (2023) tested a broader space of preprocessing-classifier combinations (time-domain, WPD, CWT, signal-to-image conversion) on induction-motor bearing faults under noise and load variation, finding that LeNet-5 with CWT preprocessing was the most robust pair while signal-to-image conversion inputs struggled in noisy conditions (Moysidis et al., 2023).

9. Applications

Vibration analysis has spread well beyond its traditional home in rotating-machinery condition monitoring. The application papers gathered for this review fall into four broad classes: rotating machinery and motors, structural health monitoring of civil and aerospace structures, manufacturing process monitoring, and a long tail of more specialised deployments in agriculture, mining, rail, and fluid power. The dominant trend across all four classes is the convergence on a common technology stack: an accelerometer (often MEMS), a time-frequency representation of the signal, and a learned classifier on top.

9.1. Rotating machinery and motors

The single largest cluster of papers in this collection targets rotating machinery, bearings, gearboxes, and motors. Khalil and Rostam (2024) built a semi-automated predictive-maintenance system for induction-draft centrifugal fans in cement plants using a 100 mV/g accelerometer and FFT energy features mapped to bearing degradation stages under extreme multi-axis loading (Khalil & Rostam, 2024). Ayankoso et al. (2024) showed that for induction motors, vibration signals reached near 100% accuracy on bearing faults and misalignments, while current signals trailed at 87.41%, confirming vibration as the better diagnostic medium for mechanical defects (Ayankoso et al., 2024). Wu et al. (2025) addressed a different bottleneck — the cost of acquiring hardware data — by generating virtual 3-axis sensor data for Y2-132M-4 motors and training a random-forest classifier that reached 92% on simulated mechanical and electrical faults (Wu et al., 2025).

Daga and Garibaldi (2019) handled the harder problem of aerospace high-speed bearings, where confounding operational variables typically blur fault signatures. Their hypothesis-testing pipeline using PCA and Mahalanobis-distance novelty indices produces a lossless dimensionality-reduced score that suppresses operational variability without requiring human supervision (Daga & Garibaldi, 2019). Cui et al. (2025) targeted railway-EMU motor bearings, fusing variational mode decomposition with a CNN-BiLSTM-ResNet hybrid and a one-class SVM. Their ground-based monitoring system reached 95.2% accuracy on highly coupled non-stationary signals and provided early warning for previously unseen fault states (Cui et al., 2025). Pavithra and Ramachandran (2025) closed the loop on CNC and rolling-bearing applications with their EMD-IMF DCNN that retained 100% classification accuracy under additive Gaussian noise up to 10 dB (Pavithra & Ramachandran, 2025).

9.2. Structural health monitoring

SHM applications are split between civil infrastructure and aerospace structures, and each subdomain has unique constraints. Dragos et al. (2024) tackled the data-synchronisation problem in wireless SHM networks on full-scale road bridges. Independent clock drift can produce time lags of up to 300 ms across

nodes, and any lag longer than 10 ms is enough to drop the Modal Assurance Criterion (MAC) below 0.9 and degrade modal-shape integrity. Their cross-spectral density phase-angle method estimated the lags accurately enough to keep MAC values intact (Dragos et al., 2024).

In aerospace and energy SHM, Garcia and Tcherniak (2019) demonstrated that a minimal setup of one accelerometer and one electromechanical actuator can detect and track trailing-edge damage progression on wind-turbine blades, a result that matters for the economics of large-scale blade monitoring (Garcia & Tcherniak, 2019). Puruncajas et al. (2020) extended SHM to offshore wind-turbine jacket foundations using triaxial piezoelectric accelerometers feeding raw signals into vibration-response-only CNNs (Puruncajas et al., 2020). Gómez Muñoz et al. (2025) closed the gap between expensive measurement and cheap deployment by training a GoogLeNet on CWT scalograms from a laser Doppler vibrometer and transferring the learned features to accelerometer data on metallic plates, achieving near-100% accuracy without re-training on the cheap-sensor data (Gómez Muñoz et al., 2025).

9.3. Manufacturing process monitoring

Vibration sensing in manufacturing is increasingly competing with acoustic monitoring for tool-state estimation. Han et al. (2021) ran two scenarios on a CNC milling testbed: a classification task on fault states (audio signals only) and a regression task for estimating machine settings — depth of cut, width of cut, feed rate, spindle speed — in which audio and vibration signals were compared using FFT features reduced by PCA before being fed to NNR and SVR models. Vibration came out more accurate than audio for the regression task on continuous machine parameters (Han et al., 2021). The SNR advantage of contact vibration sensing is large enough to justify the installation cost in production settings even though microphone arrays are usually cheaper to deploy.

9.4. Niche and emerging deployments

A long tail of application papers shows vibration sensing entering domains where it was previously absent or marginal. Al-Haddad et al. (2024) instrumented multirotor UAVs with ADXL335 accelerometers and showed that DWT multi-resolution analysis significantly outperforms Kalman filtering for fault detection. The Kalman filter, applied to a 100–200 Hz window, missed a low-frequency fault signature at approximately 51 Hz that the DWT pipeline recovered (Al-Haddad et al., 2024). Mystkowski et al. (2022) built a rugged, low-cost monitoring prototype around a MEMS ADXL356 and a piezoelectric accelerometer to track the grease gearbox and drive shaft of an agricultural hay rotary tedder in the field, using only on-microcontroller RMS, crest-factor, and kurtosis analysis (Mystkowski et al., 2022). The simplicity is the point: predictive maintenance at this price point is now available to applications where it was previously uneconomic.

Alharbi et al. (2023) covered the opposite extreme of scale, using distributed optical fibre sensing (DOFS) together with acoustic sensors and accelerometers to monitor belt-conveyor idlers in open-pit mining operations. ANN classifiers on DOFS data reached 99% accuracy on conveyor-idler damage in real time, which is a sensing density that would be impossible with discrete accelerometers (Alharbi et al., 2023). Li et al. (2025) demonstrated multi-modal fusion in industrial fluid power: combining high-frequency pressure sensors with ultra-low-frequency vibration sensors (0–0.5 Hz) enabled tracking of hydraulic-valve degradation and remaining useful life using FFT, WPD, and time-domain features (Li et al., 2025).

10. Open challenges and future directions

The current state of the field is healthy by most measures. Sensor performance is improving across families, machine-learning pipelines deliver consistent >95% classification accuracy on canonical benchmarks, and the cost of useful instrumentation has dropped enough to make predictive maintenance economically viable in domains that were previously out of reach. The remaining open problems are different from the ones that dominated the field a decade ago. They are less about transducer physics and more about integration, reproducibility, deployment economics, and the credibility of machine-learning results.

10.1. Reproducibility and reporting standards

The most pressing problem is also the most concrete. Bagri et al. (2024) audited the rotating-machinery vibration-analysis pipeline and found that only 15.41% of recent papers report their preprocessing methods and only 10.15% include experimental validation. This is a serious gap: a classification accuracy reported without the preprocessing parameters is not a result a reader can reproduce. The fix is cultural rather than technical, but a similar pattern emerged in computer vision a decade ago and was partly solved by community pressure on benchmarks and code release. Vibration analysis does not yet have an equivalent of ImageNet's accuracy-versus-recipe transparency, and the field would benefit from one.

A related issue is that benchmark accuracy on canonical bearing datasets (CWRU, MFPT, KAIST, PU) is now saturated, with multiple methods reaching 99% or 100% (Verstraete et al., 2017; Chennana et al., 2025; LGAF-Swin Transformer, 2025). Further accuracy improvements on these datasets are no longer informative. What the field needs is harder evaluation: cross-machine generalisation, transfer between operating conditions, and accuracy under realistic noise and occlusion. The transfer-learning result of Gómez Muñoz et al. (2025), where a model trained on laser-vibrometer data successfully classified accelerometer data, is the kind of evaluation that should become standard.

10.2. Sensor fusion

Multi-modal fusion is now a clear research frontier. Han et al. (2021) found that microphone and vibration signals carry overlapping but non-identical diagnostic information for CNC monitoring, with vibration delivering the more accurate result; combining them would presumably yield more than either alone. Li et al. (2025) demonstrated this directly by fusing high-frequency pressure sensors with ultra-low-frequency vibration data for hydraulic-valve degradation tracking. Ayankoso et al. (2024) compared vibration to current signals on induction motors and found vibration superior at near 100% versus 87.41%, but the obvious experiment of fusing the two remains under-reported in this collection. The barrier is rarely the sensor hardware. It is the synchronisation, calibration, and learning pipeline that combines heterogeneous streams without introducing artefacts.

10.3. Self-powered and energy-aware sensing

Two complementary directions address the energy budget of distributed sensing. Triboelectric nanogenerators (Mehamud et al., 2022) eliminate the need for an external power supply altogether, at the cost of mechanical wear and limited high-frequency stability. Digital MEMS accelerometers in low-power mode (Parisi et al., 2022) reduce energy consumption by 40–80% depending on the transmission protocol but introduce sampling-rate-related amplitude errors. Neither approach is general yet, but both signal a real research direction: predictive-maintenance economics will improve substantially if a sensor can survive in the field for years without a battery change. The same logic applies to harvesters paired with low-power MCUs running on-device feature extraction rather than streaming raw data.

10.4. Edge AI and on-device inference

Most of the deep-learning pipelines discussed in Section 8 assume that signal processing and inference happen on dedicated GPUs. For deployments at scale, that assumption is uneconomic. Łuczak (2024) demonstrated that a CWT-CNN can process a fault classification in under 50 ms, with Grad-CAM interpretability layered on top. Mystkowski et al. (2022) showed that microcontroller-based RMS, crest-factor, and kurtosis analysis is enough for field-deployed agricultural monitoring. The gap between these two endpoints — a GPU-class CNN versus a 32-bit MCU's statistical features — is where the practical research lies. Quantised CNNs, hardware accelerators for FFT and wavelet computation, and lightweight Transformer variants all sit in this gap, and the literature in this collection only partially addresses them.

10.5. Material durability and harsh environments

For applications where the sensor will be deployed permanently on a structure or machine, durability is now the rate-limiting factor. The FECA eddy-current array (Song et al., 2021) and underwater FBG deployments (de la Torre et al.,

2021) demonstrate that surviving aggressive environments is feasible, but each case required substantial engineering. The flexible piezoresistive films of Mou et al. (2025) and the polymeric PVDF designs of Ge et al. (2023; 2024) point toward a future where the sensor is conformally integrated with the structure, but neither modality has reached the dynamic range or thermal stability of conventional MEMS yet. Bridging that gap is a materials problem rather than a transduction problem.

10.6. Explainability and operator trust

A subtler but increasingly important challenge is whether plant operators trust the outputs of black-box deep-learning models. Garcia and López (2025) note that explainable-AI tools such as SHAP and Grad-CAM are increasingly necessary to translate CNN outputs into actionable insights, and that hybrid physics-data models (such as physics-informed neural networks) help when labelled data is scarce. The same paper identifies concept drift — slow changes in the underlying machinery distribution — as a problem that purely data-driven models handle poorly. Both points imply that the next generation of pipelines will be more transparent and more closely coupled to physical models than the current pure-CNN pipelines.

10.7. Standardization

A final challenge is the absence of standard performance reporting for sensors and pipelines. Nichani et al. (2025) propose a factorial design across measures of amplitude to quantify uncertainty, and Aiordăchioaie (2025) constructs a "global quality criterion" that ranks diagnostic methods on recognition rate, window length, response time, computational load, and algorithm complexity together. Neither has been adopted widely. The field would benefit from a small set of standard test rigs, reporting formats, and uncertainty budgets analogous to the metrology standards that govern other measurement domains.

Conclusion

Vibration sensing has matured from a single-discipline measurement technique into a tightly coupled stack of sensors, signal processing, and machine-learning models. The 88 papers reviewed here collectively support three conclusions.

First, no sensor family dominates across applications. Electrodynamic geophones own the sub-hertz noise floor; capacitive and piezoelectric MEMS deliver the best cost-to-performance ratio for predictive maintenance and SHM; laser Doppler vibrometry and Fabry-Pérot interferometers occupy the precision and non-contact niches; FBG sensors are the modality of choice for distributed and harsh-environment sensing; and a handful of emerging modalities — triboelectric, flexible piezoresistive, eddy-current arrays — fit application

envelopes that are too narrow for the mainstream device families. The right comparator is always the specific application, not the sensor in isolation.

Second, the boundary between signal processing and machine learning has effectively dissolved. Time-frequency representations are now produced primarily so that convolutional networks can consume them as images, and the dominant pipeline (CWT scalogram → pre-trained CNN) reaches >99% accuracy on benchmark bearing datasets routinely enough that benchmark accuracy is no longer the differentiating metric. The frontier has moved to transfer learning across sensor types, attention-based architectures, generative-adversarial data augmentation, and explainability tools that let an engineer audit a prediction.

Third, the remaining open problems are not transducer-physics problems. They are reporting standards, reproducibility, multi-modal fusion, edge inference, durability under harsh environments, and operator trust. The field has acquired both the sensors and the algorithms it needs; the next phase of progress will come from integrating them at scale with the discipline that production deployments demand. The literature already contains several models of how to do this well — the wireless SHM synchronisation work of Dragos et al. (2024), the transfer-learning result of Gómez Muñoz et al. (2025), the systematic audit of Bagri et al. (2024) — and the task ahead is to make these practices routine rather than exemplary.

References

[1] Aiordăchioaie, D. (2025). A comparative analysis of fault detection and process diagnosis methods based on a signal processing paradigm. Discover Applied Sciences, 7(10), 10. https://doi.org/10.1007/s42452-024-06390-3

[2] Al-Haddad, L. A., et al. (2024). Vibration Signal Processing for Multirotor UAVs Fault Diagnosis: Filtering or Multiresolution Analysis?. Eksploatacja i Niezawodność – Maintenance and Reliability, 26(1). 10.17531/ein/176318

[3] Alharbi, F., et al. (2023). A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models. Sensors, 23, 1902.

[4] Arts, L. P. A., & van den Broek, E. L. (2022). The fast continuous wavelet transformation (fCWT): real-time, high quality, noise-resistant, time-frequency analysis. Nature Computational Science, 2, 47-58. https://doi.org/10.1038/s43588-021-00183-z

[5] Ayankoso, S., et al. (2024). Performance of vibration and current signals in the fault diagnosis of induction motors using deep. Structural Health Monitoring, 25(1), 211-212.

[6] Bagri, I., Tahiry, K., Hraiba, A., et al. (2024). Vibration Signal Analysis for Intelligent Rotating Machinery Diagnosis and Prognosis: A Comprehensive Systematic Literature Review. Vibration, 7(4), 1013-1062. https://doi.org/10.3390/vibration7040054

[7] Bointon, P., Todhunter, L., Clare, A., et al. (2020). Performance Verification of a Flexible Vibration Monitoring System. Machines, 8, 3.

[8] Bonopera, M. (2022). Fiber-Bragg-Grating-Based Displacement Sensors: Review of Recent Advances. Materials, 15(16), 5561. https://doi.org/10.3390/ma15165561

[9] Chennana, A., et al. (2025). s41598-025-93133-y. Scientific Reports. 10.1038/s41598-025-93133-y

[10] Cui, Y., Zhang, W., & Wang, Z. (2025). Abnormal Vibration Signal Detection of EMU Motor Bearings Based on VMD and Deep Learning. Sensors, 25, 5733. 10.3390/s25185733

[11] Daga, A. P. (2019). Vibration Monitoring: Gearbox Identification and Faults Detection. Information, 10, 204.

[12] Dalla Vedova, M. D. L., Quattrocchi, G., Aimasso, A., et al. (2024). Rapid prototyping of FBG-based optical sensors for vibration analysis of mechatronic systems. Journal of Physics: Conference Series, 2698(1), 012004. https://doi.org/10.1088/1742-6596/2698/1/012004

[13] Darwish, A., Halkon, B., Rothberg, S., et al. (2022). A comparison of time and frequency domain-based approaches to laser Doppler vibrometer instrument vibration correction. Journal of Sound and Vibration, 520, 116607. https://doi.org/10.1016/j.jsv.2021.116607

[14] Darwish, A., Halkon, B., Oberst, S., et al. (2020). Correction of laser doppler vibrometer measurements affected by sensor head vibration using time domain techniques. EASD Procedia EURODYN, 2, 4842-4850. https://doi.org/10.47964/1120.9392.20444

[15] Dragos, K., et al. (2024). Frequency-domain synchronization of structural health monitoring data. Journal of Sound and Vibration, 571, 118017.

[16] Faysal, A., et al. (2021). Noise Eliminated Ensemble Empirical Mode Decomposition Scalogram Analysis for Rotating Machinery Fault Diagnosis. Sensors, 21, 8114. 10.3390/s21238114

[17] Gangopadhyay, T. K. (2004). Prospects for Fibre Bragg Gratings and Fabry-Perot Interferometers in fibre-optic vibration sensing. Sensors and Actuators A: Physical, 113, 20-38.

[18] Garcia, J., Rios-Colque, L., Peña, A., et al. (2025). Condition Monitoring and Predictive Maintenance in Industrial Equipment: An NLP-Assisted Review of Signal Processing, Hybrid Models, and Implementation Challenges. Applied Sciences, 15, 5465. https://doi.org/10.3390/app15105465

[19] Garcia, D., & Tcherniak, D. (2019). An experimental study on the data driven structural health monitoring. Mechanical Systems and Signal Processing.

[20] Ge, C., & Cretu, E. (2023). A polymeric piezoelectric MEMS accelerometer with high sensitivity, low noise density, and an innovative manufacturing approach. Microsystems & Nanoengineering, 9, 151. 10.1038/s41378-023-00628-7

[21] Ge, C., & Cretu, E. (2024). Polymeric piezoelectric accelerometers with high sensitivity, broad bandwidth, and low noise density for organic electronics and wearable microsystems. Microsystems & Nanoengineering, 10, 61. 10.1038/s41378-024-00704-6

[22] Gong, X., Kuo, Y.-C., Zhou, G., et al. (2023). An aerosol deposition based MEMS piezoelectric accelerometer for low noise measurement. Microsystems & Nanoengineering, 9, 23. 10.1038/s41378-023-00484-5

[23] Goyal, D., & Pabla, B. S. (2016). The vibration monitoring methods and signal processing techniques for structural health monitoring: a review. Archives of Computational Methods in Engineering, 23(4), 585-594. 10.1007/s11831-015-9145-0

[24] Gómez Muñoz, C. Q., et al. (2025). Structural health monitoring through deep learning: a study on scalogram-based vibration signal. Journal of Low Frequency Noise, Vibration and Active Control, 45(1), 418-419.

[25] Han, S., Mannan, N., Stein, D. C., Pattipati, K. R., & Bollas, G. M. (2021). Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems. Manuscript.

[26] Hassan, I. U., Panduru, K., & Walsh, J. (2024). An In-Depth Study of Vibration Sensors for Condition Monitoring. Sensors, 24(3), 740. https://doi.org/10.3390/s24030740

[27] Huseynzade, E., & Abdullayev, H. (2024). Usefulness of Vibration Analysis Techniques And Sensors to Improve the Monitoring of Industrial Equipment. Journal of Innovations in Business and Industry, 02(04), 229-234. 10.61552/JIBI.2024.04.004

[28] Jasiński, J., Pluta, M., Trojanowski, R., et al. (2025). Performance of Acoustic, Electro-Acoustic and Optical Sensors in Precise Waveform Analysis of a Plucked and Struck Guitar String. Sensors, 25(21), 6514. https://doi.org/10.3390/s25216514

[29] Kalybek, M., Bocian, M., & Nikitas, N. (2021). Performance of Optical Structural Vibration Monitoring Systems in Experimental Modal Analysis. Sensors, 21(4), 1239. https://doi.org/10.3390/s21041239

[30] Kedadouche, M., Thomas, M., & Tahan, A. (2016). A comparative study between Empirical Wavelet Transforms and Empirical Mode Decomposition Methods: Application to bearing defect diagnosis. Mechanical Systems and Signal Processing, 81, 88–107.

[31] Khalil, A. F. (2024). Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study. Engineering, Technology & Applied Science Research, 14(2), 13181-13189.

[32] Kirchhoff, R., et al. (2018). Huddle test measurement of a near Johnson noise limited geophone.

[33] Kolar, D., Lisjak, D., Pająk, M., & Pavković, D. (2020). Fault diagnosis of rotary machines using deep convolutional neural network with wide three axis vibration signal input. Sensors, 20(14), 4017.

[34] Lamberti, A., De Pauw, B., & Vanlanduit, S. (2016). Development of an Optical Fiber Sensor Interrogation System for Vibration Analysis. Journal of Sensors, 2016, 5204581. https://doi.org/10.1155/2016/5204581

[35] Lamberti, A., Chiesura, G., Luyckx, G., et al. (2015). Dynamic Strain Measurements on Automotive and Aeronautic Composite Components by

Means of Embedded Fiber Bragg Grating Sensors. Sensors, 15, 27174-27200. https://doi.org/10.3390/s151027174

[36] Li, E., Jian, J., Yang, F., et al. (2024). Characterization of Sensitivity of Time Domain MEMS Accelerometer. Micromachines, 15(2), 227. 10.3390/mi15020227

[37] Li, X., et al. (2025). Multi-Sensor Data Fusion and Vibro-Acoustic Feature Engineering for Health Monitoring and Remaining Useful Life Prediction of Hydraulic Valves. Sensors, 25, 6294.

[38] Lupea, I., & Lupea, M. (2025). Continuous Wavelet Transform and CNN for Fault Detection in a Helical Gearbox. Applied Sciences, 15, 950. 10.3390/app15020950

[39] Mehamud, I., et al. (2022). Machine condition monitoring enabled by broad range vibration frequency detecting triboelectric nano-generator (TENG)-based vibration sensors. Nano Energy, 98, 107292. https://doi.org/10.1016/j.nanoen.2022.107292

[40] Miranda, V. R., & Landre Jr., J. (2018). Comparison of the signal characteristics measured by a MEMS and a Piezoelectric accelerometers. International Journal of Advanced Engineering Research and Science (IJAERS), 5(11), 21. 10.22161/ijaers.5.11.21

[41] Mohd Ghazali, M. H., & Rahiman, W. (2021). Vibration Analysis for Machine Monitoring and Diagnosis: A Systematic Review. Shock and Vibration, 2021, 1-25. https://doi.org/10.1155/2021/9469318

[42] Mou, Y., et al. (2025). Flexible Vibration Sensors with Omnidirectional Sensing Enabled by Femtosecond Laser-Assisted Fabrication. Polymers, 17(2), 211. https://doi.org/10.3390/polym17020211

[43] Moysidis, D. A., et al. (2023). A Study of Noise Effect in Electrical Machines Bearing Fault Detection and Diagnosis considering Different Representative Feature Models. Machines, 11, 1029.

[44] Mystkowski, A., et al. (2022). Design and Evaluation of Low-Cost Vibration-Based Machine Monitoring System for Hay Rotary Tedder. Sensors, 22, 4072.

[45] Nichani, K., Martin, V. S., Jaelani, Y., et al. (2025). On suitability of measures of amplitude for quality assurance of vibration sensing systems in structural health monitoring. Structural Concrete, 26(5), 5619-5637. https://doi.org/10.1002/suco.70070

[46] Ompusunggu, A. P., et al. (2021). Long-term condition monitoring of critical assets with low-cost MEMS accelerometers. Procedia CIRP, 104, 1389–1394. 10.1016/j.procir.2021.11.234

[47] Pacheco-Chérrez, J., et al. (2022). 1-s2.0-S1350630722004897-main. Engineering Failure Analysis, 139, 106515.

[48] Parisi, E., Moallemi, A., Barchi, F., et al. (2022). Time and Frequency Domain Assessment of Low-Power MEMS Accelerometers for Structural Health Monitoring. IEEE, 10.1109/MetroInd4.0IoT54413.2022.9831707

[49] Pavithra, R., & Ramachandran, P. (2025). Predictive machine health monitoring using deep convolution neural network for noisy vibration signal of rotating machine using empirical mode decomposition. Discover Applied Sciences, 7, 247. https://doi.org/10.1007/s42452-025-06650-w

[50] Peng, Z. K., Tse, P. W., & Chu, F. L. (2005). A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. Mechanical Systems and Signal Processing, 19, 974–988. https://doi.org/10.1016/j.ymssp.2004.01.006

[51] Puruncajas, B., Vidal, Y., & Tutivén, C. (2020). Vibration-response-only structural health monitoring for offshore wind turbine jacket foundations via convolutional neural networks. Sensors, 20(12), 3429. https://doi.org/10.3390/s20123429

[52] Quattrocchi, G., Berri, P. C., Dalla Vedova, M. D. L., et al. (2021). Optical fibers applied to aerospace systems prognostics: design and development of new FBG-based vibration sensors. IOP Conference Series: Materials Science and Engineering, 1024(1), 012095. https://doi.org/10.1088/1757-899X/1024/1/012095

[53] Quiles-Cucarella, E., et al. (2025). Optimizing Bearing Fault Diagnosis in Rotating Electrical Machines Using Deep Learning and Frequency Domain Features. Applied Sciences, 15, 3132. 10.3390/app15063132

[54] Rai, V. K., & Mohanty, A. R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. Mechanical Systems and Signal Processing, 21(6), 2607–2615. https://doi.org/10.1016/j.ymssp.2006.12.004

[55] Rodrigues, J. V. O., Pedroso, M. P. G., Silva, F. F. B., et al. (2021). Performance evaluation of accelerometers ADXL345 and MPU6050 exposed to random vibrational input. Research, Society and Development, 10(15), e286101523082. 10.33448/rsd-v10i15.23082

[56] Rodríguez García, Y., Corres, J. M., & Goicoechea, J. (2010). Vibration Detection Using Optical Fiber Sensors. Journal of Sensors, 2010, 936487. https://doi.org/10.1155/2010/936487

[57] Shen, P., Gao, C., Lu, M., et al. (2025). A comparative study of non-contact acoustic emission monitoring using laser doppler vibrometer in CFRP

tensile testing. Journal of Physics: Conference Series, 2954(1), 012084. https://doi.org/10.1088/1742-6596/2954/1/012084

[58] Silva, L. L., Rodrigues, J. R., Passaro, A., et al. (2017). Dynamic Sensitivity Study of MEMS Capacitive Acceleration Transducer Based on Analytical Squeeze Film Damping and Mechanical Thermoelasticity Approaches. arXiv.

[59] Song, Y., et al. (2021). The durability of flexible eddy current array (FECA) sensors in harsh service environments. Scientific Reports.

[60] Sony, S., Laventure, S., & Sadhu, A. (2019). A literature review of next-generation smart sensing technology in structural health monitoring. Structural Control and Health Monitoring, 26, e2321. https://doi.org/10.1002/stc.2321

[61] Symes, R. W., Varley, R. J., St John, N., et al. (2025). Comparative analysis of MEMS (Micro-Electro-Mechanical Sensor) and IEPE (Integrated Electronics Piezo-Electric) accelerometers for measurement of wide spectrum damping. Measurement, 242, 115963. 10.1016/j.measurement.2024.115963

[62] Taghavi, M., Abedi, A., Parsanasab, G.-M., et al. (2022). Closed-Loop MOEMS Accelerometer. Optica Publishing Group.

[63] Tama, B. A., et al. (2022). Recent advances in the application of deep learning for fault. Artificial Intelligence Review, 4670-4709. 10.1007/s10462-022-10293-3

[64] Tsai, J.-M., Sun, I.-C., & Chen, K.-S. (2021). Realization and performance evaluation of a machine tool vibration monitoring module by multiple MEMS accelerometer integrations. International Journal of Advanced Manufacturing Technology, 114, 465-479. 10.1007/s00170-021-06856-3

[65] Velázquez-Carreón, F., Guo, K., Wang, H., et al. (2026). Statistical modeling and performance characterization of bridge-type FBG sensors for low-frequency vibration measurement. Measurement, 266, 120435. https://doi.org/10.1016/j.measurement.2026.120435

[66] Verstraete, D., et al. (2017). Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of. Shock and Vibration, 2017. 10.1155/2017/5067651

[67] Wang, H., et al. (2025). Fault Diagnosis of Rolling Bearing Acoustic Signal Under Strong Noise Based on WAA-FMD and LGAF-Swin Transformer. Processes, 13, 2742. 10.3390/pr13092742

[68] Wei, et al. (2025). WEI,+6-V15N3(2025)-IJETI14774(314-331). International Journal of Engineering and Technology Innovation, 15(3), 314-331.

[69] Wu, C., Doraisamy, V., & Rasalingam, R. R. (2025). A digital transformation framework for real-time vibration monitoring using virtual sensors and machine learning. paperASIA, 41(4b), 460–468. https://doi.org/10.59953/paperasia.v41i4b.601

[70] Xiong, R., Xu, X., Liu, Y., et al. (2025). A miniaturized MEMS accelerometer with anti-spring mechanism for enhancing sensitivity. Microsystems & Nanoengineering, 11, 42. 10.1038/s41378-024-00826-x

[71] Xu, Y., Yang, Y., & Wu, Y. (2020). Eddy Current Testing of Metal Cracks Using Spin Hall Magnetoresistance Sensor and Machine Learning.

[72] Yu, M., Wu, X., Zhao, L., et al. (2025). Monolithically integrated triaxial high-performance high-g accelerometer for high shock vibration signal measurements. Microsystems & Nanoengineering, 11, 206. 10.1038/s41378-025-01051-w

[73] Yu, M., Zhao, L., Chen, S., et al. (2023). Monolithically integrated triaxial high-performance micro accelerometers with position-independent pure axial stressed piezoresistive beams. Microsystems & Nanoengineering, 9, 9. 10.1038/s41378-022-00474-z

[74] Zhang, C., Mousavi, A. A., Masri, S. F., et al. (2022). Vibration feature extraction using signal processing techniques for structural health monitoring: A review. Mechanical Systems and Signal Processing, 177, 109175. https://doi.org/10.1016/j.ymssp.2022.109175

[75] Zhang, S., et al. (2022). State of the art on vibration signal processing towards data-driven gear fault diagnosis. IET Collaborative Intelligent Manufacturing, 4(4), 249-266. 10.1049/cim2.12064

[76] Zhou, Y., Ma, Z., & Fu, L. (2025). A Review of Key Signal Processing Techniques for Structural Health Monitoring: Highlighting Non-Parametric Time-Frequency Analysis, Adaptive Decomposition, and Deconvolution. Algorithms, 18, 318. 10.3390/a18060318

[77] de la Torre, O., Floris, I., Sales, S., et al. (2021). Fiber Bragg Grating Sensors for Underwater Vibration Measurement: Potential Hydropower Applications. Sensors, 21(13), 4272. https://doi.org/10.3390/s21134272

[78] Łuczak, D. (2024). Machine Fault Diagnosis through Vibration Analysis: Continuous Wavelet Transform with Complex Morlet Wavelet and Time–Frequency RGB Image Recognition via Convolutional Neural Network. Electronics, 13, 452. https://doi.org/10.3390/electronics13020452