Research Article

A Data Fusion Method Combining Image, Sensor, and Survey Data for Efficiency and Usability Analysis of Electric Power Tools in Industrial Environments

Kader NİKBAY OYLUM[1*], Turgay Tugay BİLGİN[2]

[1] Trex Dijital Akıllı Üretim Sistemleri A.Ş., Orcid ID: https://orcid.org/0000-0002-5218-9218, kaderoylum@trex.com.tr
[2] Bursa Teknik Üniversitesi, Orcid ID: https://orcid.org/0000-0002-9245-5728, turgay.bilgin@btu.edu.tr
[*] Correspondence: kaderoylum@trex.com.tr; Tel.: 444 3 468

Received: 21 February 2025
Revised: 28 May 2025
Accepted: 14 June 2025
Published: 28 June 2025

Reference: Nikbay Oylum, K., & Bilgin, T. T. (2024). A data fusion method combining image, sensor, and survey data for efficiency and usability analysis of electric power tools in industrial environments. The European Journal of Research and Development, 5(1), 68-83.

Abstract

The increasing integration of advanced technologies and automation in industrial production has heightened the importance of operational efficiency and safety. Among the critical components influencing workforce performance and product quality is the effective use of electric hand tools. However, the limited availability of comprehensive datasets and the absence of robust labeling methodologies present significant challenges for accurate data analysis and predictive modeling. This study addresses these limitations by incorporating field-collected data and multiple data acquisition techniques to identify relevant features for machine learning applications. An initial dataset comprising 51 attributes was systematically reduced to 16 through feature selection processes, enhancing its suitability for subsequent computational modeling. Several classification algorithms were evaluated for data labeling, with the Decision Tree method demonstrating superior performance in terms of accuracy. Despite these promising results, the dataset's limited sample size (64 individuals) restricts the generalizability and reliability of machine learning outcomes. To mitigate this constraint, data augmentation techniques will be employed to generate synthetic instances, thereby expanding the dataset. Upon achieving a sufficient sample size, machine learning models will be developed to predict individuals' proficiency with electric hand *tools. This research contributes to the foundational knowledge required for efficient data collection, accurate labeling, and the development of predictive models in industrial settings.*

Keywords: Data Fusion, Machine Learning, Digitization, Data Labeling, Predictive Modeling

1. Introduction

The efficiency and ergonomic use of electric hand tools in industrial production processes are critical factors influencing workforce productivity, product quality, and occupational health and safety. Accurate and comprehensive data are essential for assessing and enhancing the performance of hand tools within manufacturing environments. However, conventional evaluation approaches typically depend on a single source of data and are limited in their capacity to holistically incorporate real-time performance monitoring, ergonomic assessments, and user feedback.

To address these limitations, data fusion presents a robust methodology that synthesizes information from multiple sources, enabling the extraction of more reliable and meaningful insights. This study explores the enhancement of efficiency and usability analysis for electric hand tools through the integration of image data, sensor data, and survey responses. Image data are utilized to monitor and evaluate hand movements of operators within industrial settings, while sensor data capture key dynamic parameters from the tools—such as acceleration, gyroscopic activity, and spatial coordinates along the x, y, and z axes. In parallel, survey data provide subjective evaluations based on user experience and perceived ergonomics. The application of data fusion to combine these three data streams yields a more comprehensive and accurate analysis than could be achieved by considering each source in isolation.

Furthermore, the integration of data from heterogeneous sources not only enhances data diversity within industrial processes but also reduces uncertainty, thereby supporting more robust and reliable decision-making. The real-time processing of sensor and image data facilitates instantaneous monitoring of tool operations and operator interactions. This capability enables significant operational improvements, including predictive maintenance, energy consumption optimization, and increased workforce productivity. Moreover, incorporating subjective feedback from operators into the analysis contributes to enhanced user experiences and informs ergonomically optimized tool design through data-driven insights.

Despite these advantages, the practical implementation of data fusion in industrial settings presents several technical challenges. These include the synchronization of heterogeneous data streams, the computational demands of processing large volumes of data, the requirement for real-time analytical capabilities, and concerns related to data security and privacy. To address these issues, this study proposes a data fusion-based framework for evaluating the performance and ergonomic suitability of electric hand tools. The paper details the data acquisition protocols, preprocessing strategies, and the technical implementation of the fusion methodology.

The remainder of this paper is structured as follows: Section 2 presents a review of the relevant literature on data fusion techniques; Section 3 outlines the proposed methodology; Section 4 discusses the experimental setup, results, and encountered technical challenges; Section 5 offers a discussion of the findings; and Section 6 concludes the study with final remarks and directions for future research.

2. Literature

In the era of big data, data fusion has emerged as a pivotal technique in the fields of analytics, machine learning, and artificial intelligence. By integrating heterogeneous data sources, data fusion enhances the accuracy, completeness, and reliability of information, thereby improving decision-support systems. This multifaceted approach is widely adopted across diverse sectors, including industrial automation, healthcare, energy management, finance, and environmental monitoring. Its growing relevance reflects the need for more holistic data interpretation frameworks capable of addressing complex, data-rich environments.

This section provides a comprehensive review of the current literature on data fusion methodologies and explores their applications across various domains, with a particular focus on their potential and challenges within industrial settings.

2.1. Data Fusion and Its Applications

Fundamental approaches to data fusion are generally categorized into early, late, and hybrid fusion methods [1]. These strategies differ based on the stage at which data integration occurs within the processing pipeline. Data fusion models have been applied across various disciplines, often contextualized within established frameworks such as the Joint Directors of Laboratories (JDL) model and Dasarathy's input-output classification scheme. Notably, advanced techniques such as Lie algebra-based multisensor data fusion have been proposed to improve data accuracy and reliability by leveraging the geometric properties of data distributions on Riemannian manifolds [2].

Recent advancements have increasingly incorporated machine learning and deep learning into data fusion processes. For instance, Q-learning-based cascade classifiers and softmax models have demonstrated effectiveness in optimizing network energy consumption while minimizing error rates in big data fusion scenarios [3]. Furthermore, attention-based mechanisms—such as the Attention-Guided Adaptive Temporal-Spatial (AGATS) model—have been successfully employed to integrate multi-source datasets in financial applications, enhancing the accuracy of stock market predictions [4].

2.2. Sensor and IoT-based Data Fusion

The integration of sensor data has emerged as a critical area of focus across a range of industries, aiming to enhance data accuracy and operational efficiency. Multi-source sensor fusion (MSDF) techniques have proven particularly effective in domains such as food safety, where they support contaminant detection through advanced methods including hyperspectral imaging, Raman spectroscopy, and chromatographic analysis [5]. In structural engineering, the fusion of computer vision and accelerometer data has led to significant improvements in dynamic displacement estimation, with notable applications in earthquake engineering and infrastructure monitoring [6].

In the context of the Internet of Things (IoT), multi-sensor data fusion plays a pivotal role in integrating heterogeneous data streams. This integration facilitates intelligent decision-making in energy management, transportation systems, and broader smart city applications [7]. Furthermore, sensor fusion techniques have been employed in Space Traffic Management (STM) to improve the precision of orbit determination, thereby reducing the likelihood of in-orbit collisions and contributing to safer space operations [8].

2.3. Data Fusion in Healthcare

The healthcare sector stands at the forefront of data fusion applications, leveraging multimodal integration to enhance diagnostic accuracy, treatment planning, and patient monitoring. For instance, in the context of COVID-19, data fusion techniques that combine clinical records, laboratory data, and radiomic features have significantly improved the prediction of hospitalization outcomes in patient cohorts [9]. In pharmaceutical research, models such as TransCDR utilize multimodal data fusion to predict cancer drug efficacy by analyzing interactions between drugs and cancer cell lines, thereby facilitating more targeted and effective therapies [10].

Furthermore, advanced multi-source data fusion approaches based on generalized data representations have shown promise in generating reliable diagnostic outcomes, particularly when data availability is limited. Quantum-inspired collaborative fusion techniques, for example, have demonstrated notable improvements in diagnostic accuracy in applications such as pneumonia detection [11]. These developments underscore the growing role of sophisticated data fusion strategies in advancing healthcare analytics and decision-making processes.

2.4. Data Fusion in Environmental and Energy Management

Data fusion techniques have also made significant contributions in the fields of energy modeling and environmental monitoring, supporting more informed policy-making and sustainable urban planning. Probabilistic data fusion algorithms based on maximum likelihood estimation have been effectively applied to model residential energy consumption patterns using comprehensive datasets such as the Residential Energy Consumption Survey (RECS), offering valuable insights for energy policy development [12]. In renewable energy applications, multi-source data fusion models have improved the accuracy and stability of power output forecasts in wind farms by employing Long Short-Term Memory (LSTM) encoder-decoder architectures combined with self-attention mechanisms [13].

In the realm of environmental management, data fusion methods that integrate citizen science data with atmospheric dispersion models have enhanced the resolution and accuracy of air pollution mapping at the city scale [14]. Furthermore, the fusion of satellite data from platforms such as Sentinel-3 SLSTR and Himawari-9 AHI has enabled high-resolution monitoring of snow cover dynamics in mountainous regions, significantly improving the temporal and spatial precision of snow distribution assessments [15].

2.5. Data Fusion in Industry and Manufacturing Processes

Within industrial processes, data fusion serves as a pivotal tool for optimizing operational efficiency and enhancing decision-making. In the context of alumina evaporation, multivariate data fusion approaches utilizing Gaussian filter-based adaptive noise reduction algorithms have been instrumental in improving production stability and process control [16]. Similarly, in the domain of vehicle maintenance, the integration of historical maintenance records with fundamental vehicle data through data fusion techniques has led to improved predictive accuracy for maintenance scheduling and failure detection [17].

Furthermore, data fusion methodologies based on strain energy density have demonstrated significant improvements in the accuracy of fatigue life predictions. By combining data from torsion fatigue tests with finite element method (FEM) simulations, these approaches have proven particularly effective even when working with limited sample sizes [18]. Such advancements underscore the versatility and impact of data fusion across a range of industrial applications.

2.6. Data Fusion and Future Perspectives

Recent studies underscore the expansive applicability of data fusion, ranging from decision support systems to the development of advanced artificial intelligence models [19]. Cross-domain data fusion—integrating heterogeneous sources such as geographic information, traffic data, social media inputs, and environmental metrics—has gained particular relevance in urban computing, where complex, data-rich environments demand integrated analytical frameworks [20]. Looking ahead, the integration of large language models (LLMs) and generative artificial intelligence is expected to further enhance the effectiveness and adaptability of data fusion techniques, enabling more sophisticated and context-aware applications.

Despite these advancements, several critical challenges remain. Key among them are the need for robust privacy protection mechanisms, the development and adoption of standardized data formats and protocols, and improvements in the computational efficiency of fusion models. Addressing these issues will be essential for ensuring the scalability, trustworthiness, and real-world applicability of data fusion technologies in both industrial and societal contexts.

3. Materials and Methods

This section provides a comprehensive overview of the data collection procedures from electric screwdriver devices, the integration of data from multiple sources, and the analytical methods employed throughout the analysis phase.

3.1. Data Collection Process

Three distinct data collection methods were employed in relation to the studied subject:

(1) Sensor Data from the Electric Screwdriver—data acquired via sensors embedded within the electric screwdriver, including acceleration and gyroscopic measurements along the X, Y, and Z axes captured by an Inertial Measurement Unit (IMU), as well as voltage and motor current readings;

(2) Hand Movement Data from Video Recordings—kinematic data on hand motions during screwdriving tasks, obtained through video recordings analyzed using Google's MediaPipe library, which tracked the X and Y axis positions of 21 key points corresponding to finger joints;

(3) Survey Data from Users—subjective responses collected through an 11-item questionnaire designed to assess user perceptions and experiences related to the use of electric screwdrivers.

Figure 1: Field Data Collection Study

Utilizing the aforementioned methods, empirical data were collected from a total of 64 participants across two distinct manufacturing facilities (Figure 1).

3.2. Ethical Approval and Informed Consent

The questionnaire employed in this study was administered as part of a doctoral dissertation project, conducted in accordance with the approval granted by the Ethics Committee of Istanbul Okan University. The development and distribution of the questionnaire adhered strictly to established ethical principles. All participants were comprehensively informed about the study's objectives, scope, and data handling procedures. Prior to participation, each individual voluntarily provided informed consent after reviewing the associated consent documentation.

Strict compliance with ethical standards was maintained throughout the research, particularly in safeguarding participants' privacy and ensuring the confidentiality of personal data. The collected data were used exclusively for scientific purposes. Accordingly, it is affirmed that all procedures were conducted in alignment with ethical research standards and institutional guidelines.

3.3. Digitization and Labeling Methodology of Survey Data

Responses obtained from the survey, aimed at determining the hand-tool usage experience of participants involved in the data collection study, were converted into an objective scoring formula (1) to be utilized in the developed models:

$$T=D(P+S+K) \tag{1}$$

The scoring formula was developed based on four primary factors:

Hand Tool Usage Experience (D): Indicates whether the participant had prior experience using similar hand tools, scored as either 0 or 1 point.

Professional Usage Requirement (P): Reflects the necessity of using hand tools as part of the participant's professional responsibilities, assigned 0 or 2 points.

Occupational Experience Duration (S): Scored on a scale from 1 to 5 points, depending on the participant's total years of industry experience.

Seniority Level (K): Evaluated from 0 to 3 points based on occupational roles, such as apprentice, operator, technician, or foreman.

Based on these four criteria, participants were classified into three experience categories: Inexperienced (0–3 points), Less Experienced (4–6 points), and Highly Experienced (7 or more points). The scoring model was operationalized through a decision tree methodology (Figure 2), and its reliability was verified through comparative analysis against manual classification outcomes.

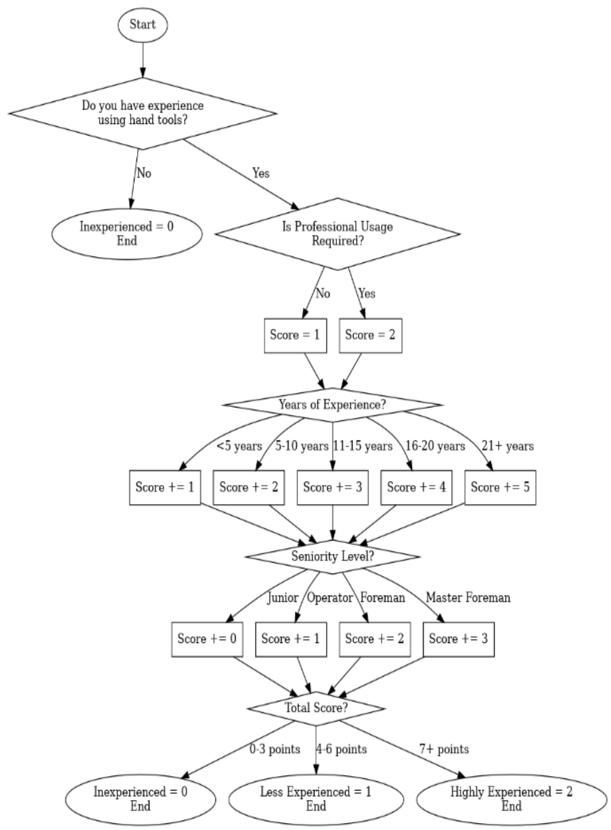

Figure 2: Labeling methodology decision tree

Based on the constructed decision tree model, the distribution of experience levels among the 64 participants from the two factories is as follows (Figure 3):

- Inexperienced: 24 participants
- Less Experienced: 21 participants
- Highly Experienced: 19 participants

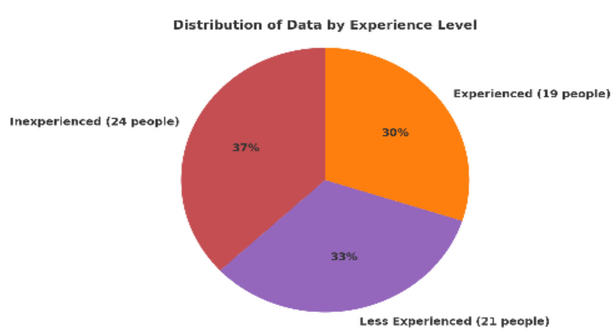

Figure 3: Experience level distribution graph

The classification conducted through the decision tree algorithm facilitated a consistent and objective evaluation of individual operator performance and experience levels. To assess the alignment with manual evaluations, the results were compared using Cohen's Kappa coefficient, which yielded a value of 0.85—indicating a high level of agreement with human raters. This validation confirms the reliability of the proposed system for potential deployment in industrial applications.

3.4. Analysis of Dataset and Derivation of Feature Relationships

The initial dataset compiled for this study comprised 51 attributes. Following a detailed analysis, it was determined that using individual position data from multiple joints on the same finger introduced redundancy. Instead, a single pair of averaged x and y coordinates per finger was deemed sufficient. Accordingly, the coordinate values of the joints on each finger were averaged, reducing the dimensionality of the dataset from 51 to 21 attributes.

In the refined dataset consisting of 21 attributes, it was observed that the significance and influence of each finger on the model varied. To explore the interrelationships among fingers, correlation analyses were performed. The results are visualized in Figure 4 as a heatmap, where the color gradient represents the strength of the correlations: dark red indicates a strong positive correlation, while dark blue indicates a strong negative correlation.

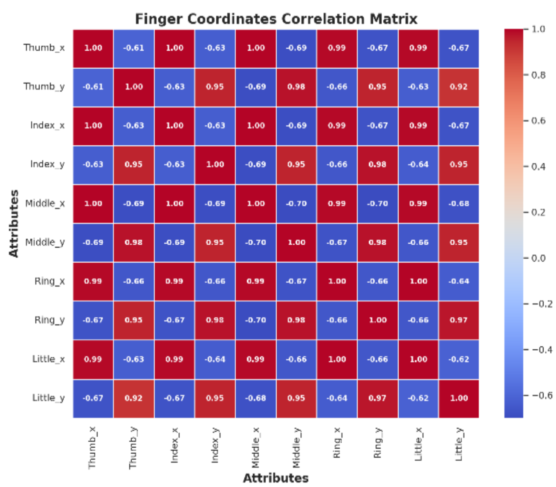

Figure 4: Finger coordinates correlation matrix

Table 1 presents the statistical properties of the X and Y coordinates for each finger, including magnitude, standard deviation, range, mean, and variance values.

Table 1: Finger movement statistics

Finger Movement Statistics				
	std	range	mean	variance

Thumb	130.761490	616.210317	1280.163710	17098.567222
Index	122.579699	614.491514	1351.283155	15025.782658
Middle	120.385253	561.494062	1306.822655	14492.609170
Ring	118.645195	505.533633	1277.258417	14076.682332
Little	118.138635	507.519428	1248.683668	13956.737136

To explore the underlying characteristics of finger movements and examine the interrelationships among them, Principal Component Analysis (PCA) was conducted. Furthermore, a uniqueness analysis was applied to identify distinct movement features associated with each finger. The results of the PCA are illustrated in Figure5.

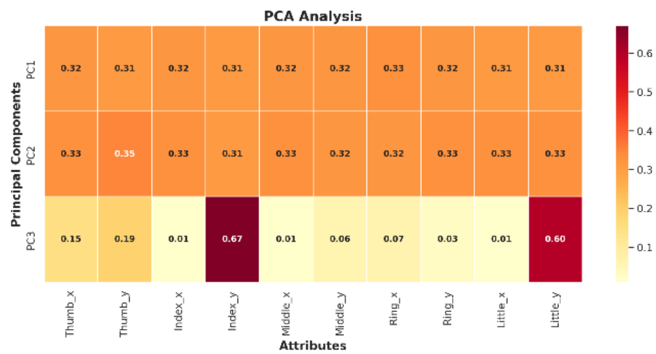

Figure 5: PCA analysis

The PCA results reveal the structural characteristics underlying finger coordination. In the first principal component, the relatively uniform loadings ranging between 0.31 and 0.32 suggest a high degree of coordination among fingers during fundamental hand movements. The second principal component displays a comparatively elevated loading for the thumb (0.35), indicating its prominent role in supporting secondary movement patterns. In the third principal component, the distinct peak associated with the index finger (0.67) underscores its critical role in executing fine motor tasks.

In addition to PCA, a uniqueness analysis was performed to assess the distinct contribution of each finger. The results of this analysis are presented in Table 2.

Table 2: Uniqueness analysis

Finger Uniqueness Scores (High = More Unique)	
Thumb	0.0172
Index	0.0202
Middle	0.0087
Ring	0.0123
Little	0.0253

An examination of the uniqueness scores reveals that the little finger exhibits the highest uniqueness value (0.0253), followed by the index finger (0.0202) and the thumb

(0.0172). In contrast, the middle finger presents the lowest uniqueness score (0.0087), with the ring finger registering the second-lowest score (0.0123).

The integrated analysis of PCA components and uniqueness scores offers valuable insights into the functional organization of hand movements. Although the little finger displays a high uniqueness score, its average contribution across the principal components suggests that it possesses an independent yet limited range of motion. In contrast, the index finger demonstrates both a high uniqueness value and a dominant contribution to the third principal component, indicating its key role in executing both independent and specialized motor tasks. Conversely, the relatively low uniqueness scores of the middle and ring fingers imply a minimal contribution to model performance, thereby suggesting the potential for dimensionality reduction through the exclusion of features associated with these fingers.

Following the analysis of inter-finger movement relationships, the study proceeded to investigate potential correlations between the VOLT and MOTOR_CURRENT variables within the dataset. Both Pearson and Spearman correlation analyses were applied to assess the strength and nature of these associations.

Table 3: Correlation analysis

Type of Correlation	Coefficient	P-Value
Pearson	-0.4555	1.926e-60
Spearman	-0.8076	1.359e-267

As presented in Table 3, the analysis results indicate a statistically significant correlation between voltage and motor current. The extremely low p-values obtained from both Pearson and Spearman tests confirm that the probability of these relationships occurring by chance is negligible, thereby supporting the robustness of the observed association.

Table 4: VIF analysis

Variable	VIF Value
VOLT	2.471645
MOTOR_CURRENT	2.471645

According to the Variance Inflation Factor (VIF) analysis presented in Table 4, the VIF values calculated for both VOLT and MOTOR_CURRENT are equal, each measured at 2.471645. This outcome satisfies the commonly accepted threshold of VIF < 5, as noted in the literature, indicating the absence of a critical multicollinearity problem. Consequently, this supports the reliability and stability of parameter estimations involving these variables in regression analyses.

Table 5: Information gain analysis

Serial Number	Attributes	MI_with_VOLT	MI_with_CURRENT
0	Wrist_x	1.418112	0.544725
1	Wrist_y	0.821802	0.339090
2	Thumb_x	1.422426	0.564319
3	Thumb_y	0.691628	0.235470
4	Index_x	1.500626	0.570305
5	Index_y	0.785272	0.245580
6	Middle_x	1.471271	0.575871
7	Middle_y	0.700679	0.192584
8	Ring_x	1.501170	0.579924
9	Ring_y	0.662587	0.215750
10	Little_x	1.553003	0.605075
11	Little_y	0.730196	0.303459
12	X_AC	0.333795	0.347479
13	Y_AC	0.196286	0.230786
14	Z_AC	0.308002	0.226366
15	X_GY	0.290699	0.343400
16	Y_GY	0.092539	0.200344
17	Z_GY	0.149488	0.185767

The information gain analysis results, as presented in Table 5, indicate that the system’s behavior is primarily influenced by the X-coordinates of hand positions. This suggests that X-coordinate features possess greater discriminatory power relative to other variables and play a pivotal role in interpreting system dynamics. Accordingly, these parameters should be given priority during the optimization of control algorithms and system modeling processes. Emphasizing the effective processing of X-coordinate data is therefore essential for enhancing the system’s responsiveness, both in terms of accuracy and overall performance.

In conclusion, the findings support the feasibility of combining the VOLT and MOTOR_CURRENT attributes into a single composite variable, POWER, without incurring significant information loss. Representing the system's energy characteristics through a unified metric offers notable advantages, including reduced computational complexity and enhanced model simplicity via dimensionality reduction. This transformation contributes to more efficient system modeling and performance optimization.

Based on the outcomes of the aforementioned analyses, the following modifications were implemented to refine the dataset:

- The attributes Middle_x, Middle_y, Ring_x, and Ring_y—identified as having minimal contribution to the model—were excluded.

• The attributes VOLT and MOTOR_CURRENT were consolidated into a newly derived attribute, POWER, computed using Equation (2):

$$POWER = VOLT.\mathrm{MOTOR\_CURRENT} \tag{2}$$

As a result of these adjustments, the original dataset comprising 51 attributes was reduced to 16 attributes, thereby enhancing its analytical efficiency and suitability for modeling.

4. Experimental Results and Technical Challenges

4.1. Experimental Results

In the development of the labeling methodology for this study, various data processing techniques were employed, including Decision Tree, Logistic Regression, Support Vector Machine (SVM), Random Forest, and Neural Network algorithms.

Table 6: Accuracy of the methods

1. Methods	Accuracy Percentage	Standard Deviation	Confidence Interval
Decision Tree	84.85	0.10	%75.83 – %93.69
Logistic Regression	69.70	0.16	%55.73 – %83.32
Support Vector Machine	66.67	0.19	%49.65 – %83.69
Random Forest	69.70	0.23	%48.90 – %90.15
Neural Network	72.73	0.20	%54.76 – %90.00

Table 6 summarizes the average accuracy, standard deviation, and confidence interval percentages for the implemented models. Among the evaluated algorithms, the Decision Tree model demonstrated the highest reliability, achieving an accuracy rate of 84.85%, accompanied by a low standard deviation and a narrow confidence interval. These metrics indicate the model's robustness and consistent performance across the dataset.

According to the results, the Neural Network method achieved the second-highest accuracy rate at 72.73%. This relatively high performance reflects the capability of deep learning models to capture and learn complex patterns within the data. Logistic Regression and Random Forest models exhibited comparable accuracy rates, each at 69.70%. In contrast, the Support Vector Machine (SVM) model recorded the lowest accuracy rate among the tested algorithms, with a value of 66.67%.

These findings provide valuable insights into the comparative performance of various classification methods applied to the dataset. Prior to finalizing the Decision Tree model—which achieved the highest accuracy—a cross-validation procedure was conducted to assess the risk of overfitting. The model yielded an average accuracy of 70% across validation folds, with noticeable performance fluctuations, suggesting potential overfitting and limitations related to the dataset's relatively small size.

To address these concerns, additional regularization techniques were employed. Specifically, Cost Complexity Pruning was applied in conjunction with Grid Search Cross-Validation (5-fold) to optimize model complexity and enhance generalization. This refined approach led to an improved accuracy of 82.38%, demonstrating a significant reduction in overfitting and increased stability of the model across different data subsets.

An additional error analysis was performed to identify conditions under which the Decision Tree model produced incorrect predictions. The analysis revealed that Features 7, 10, and 11 were significant contributors to misclassification. In particular, the model exhibited a tendency to mislabel instances when Feature7 had a value of 0. Furthermore, the low variance observed in Feature10 hindered the model's ability to generalize effectively. Similarly, elevated values in Feature11 were associated with an increased rate of prediction errors, suggesting sensitivity to this feature's distribution.

4.2. Technical Challenges

One of the primary challenges encountered in this study was the lack of an existing dataset, which necessitated the implementation of field-based data collection protocols. Securing the necessary permissions from manufacturing facilities constituted a major obstacle during the research process. Moreover, the limited availability of time for conducting on-site studies, coupled with restrictions on the number of eligible participants, further complicated the data acquisition phase.

5. Findings

According to the research findings, among the machine learning algorithms evaluated, the Decision Tree method exhibited the highest and most consistent performance, achieving an accuracy rate of 84.85% with low variance. This outcome underscores not only the theoretical effectiveness of the model but also its practical advantages for industrial applications. The inherently interpretable structure of Decision Trees supports the development of decision-support systems within production environments, enhancing their applicability in domains such as quality control, workforce planning, and maintenance strategy optimization. Furthermore, the method's low computational cost and fast prediction capabilities make it particularly well-suited for real-time industrial applications. Therefore, these results provide valuable insights into the potential for integrating the Decision Tree approach into large-scale industrial implementations.

6. Conclusions and Future Work

In this study, a series of analyses were conducted to integrate data collected through multiple field-based methods and to identify the most relevant attributes for machine learning applications. The original dataset, initially composed of 51 attributes, was reduced to 16 attributes through dimensionality reduction techniques, rendering it more suitable for the subsequent modeling stages.

As part of the data labeling methodology, several classification algorithms were evaluated, among which the Decision Tree model was selected due to its superior accuracy.

For future research, it is recognized that the current dataset—comprising data from only 64 participants—may be insufficient to fully leverage the potential of machine learning models. To address this limitation, synthetic data generation via data augmentation techniques is planned to expand the dataset. Building on this enriched dataset, a machine learning-based predictive model will be developed to estimate users' experience levels in operating hand tools, further advancing the practical applications of the proposed approach.

References

[1] Krushnasamy, V. S., & Rashinkar, P. (2017). An overview of data fusion techniques. In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 694–697). IEEE.

[2] Chen, X., Chen, C., & Lu, X. (2024). Algebraic method for multisensor data fusion. *PLOS ONE, 19*(9), e0307587.

[3] Zhang, X., Wang, J., Huang, Y., & Zhu, F. (2024). A novel industrial big data fusion method based on Q-learning and cascade classifier. *Computer Science and Information Systems, 21*(4), 1629–1649.

[4] Zhang, Q., Zhang, Y., Bao, F., Ning, Y., Zhang, C., & Liu, P. (2025). Graph-based stock prediction with multisource information and relational data fusion. *Information Sciences, 690*, 121561.

[5] Adade, S. Y.-S. S., Lin, H., Johnson, N. A. N., Nunekpeku, X., Aheto, J. H., Ekumah, J.-N., ... Chen, Q. (2025). Advanced food contaminant detection through multi-source data fusion: Strategies, applications, and future perspectives. *Trends in Food Science & Technology, 156*, 104851.

[6] Niu, Y., Li, Z., Li, J., & Sun, B. (2025). Accelerometer-assisted computer vision data fusion framework for structural dynamic displacement reconstruction. *Measurement, 242*, 116021.

[7] Çengiz, B., & Daş, R. (2022). Veri füzyonu: Veri kaynakları, mimariler, zorluklar ve çözüm yaklaşımları. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 34*(2), 899–922.

[8] Oltrogge, D., & Kerr, E. (2024). Space traffic management: Data fusion. *Acta Astronautica.*

[9] Er, A. G., Ding, D. Y., Er, B., Uzun, M., Cakmak, M., Sadee, C., ... Gevaert, O. (2024). Multimodal data fusion using sparse canonical correlation analysis and cooperative learning: A COVID-19 cohort study. * Digital Medicine, 7*, 117.

[10] Xia, X., Zhu, C., Zhong, F., & Liu, L. (2024). TransCDR: A deep learning model for enhancing the generalizability of drug activity prediction through transfer learning and multimodal data fusion. *BMC Biology, 22*, 227.

[11] Peng, Y., Zhang, W., Chen, X., Li, T., & Wang, H. (2025). Multi-source data fusion for intelligent diagnosis based on generalized representation. *Expert Systems with Applications, 228*, 121968.

[12] Bhowmik, T., Iraganaboina, N. C., & Eluru, N. (2024). A novel maximum likelihood based probabilistic behavioral data fusion algorithm for modeling residential energy consumption. *PLOS ONE, 19*(11).

[13] Anand, S., & Sonkar, S. K. (2024). Multi-source data fusion for wind farm power forecasting through AI. *Journal of the Balkan Tribological Association, 30*(3), 439–450.

[14] O'Regan, A. C., Grythe, H., Hellebust, S., Lopez-Aparicio, S., O'Dowd, C., Hamer, P. D., ... Nyhan, M. M. (2024). Data fusion for enhancing urban air quality modeling using large-scale citizen science data. *Sustainable Cities and Society, 116*, 105896.

[15] Yang, R., Zhang, Y., Wei, Q., Liu, F., & Li, K. (2024). Comparison of two data fusion methods from Sentinel-3 and Himawari-9 data for snow cover monitoring in mountainous areas. *Research in Cold and Arid Regions.*

[16] Qian, X., Xu, L., & Cui, X. (2024). Steady-state detection of evaporation process based on multivariate data fusion. *PLOS ONE, 19*(9).

[17] Chen, F., Shang, D., Zhou, G., Ye, K., & Wu, G. (2024). Multi-source data fusion for vehicle maintenance project prediction. *Future Internet, 16*, 371.

[18] Wang, C., Yao, J., Zhang, X., Wu, Y., Liu, X., Liu, H., ... Xin, J. (2024). Fatigue life data fusion method of different stress ratios based on strain energy density. *Materials, 17*, 2982.

[19] Khaleghi, B., Razavi, S. N., Khamis, A., Karray, F. O., & Kamel, M. (2009). Multisensor data fusion: Antecedents and directions. In *2009 International Conference on Signals, Circuits and Systems*. IEEE.

[20] Zou, X., Yan, Y., Hao, X., Hu, Y., Wen, H., Liu, E., ... Liang, Y. (2025). Deep learning for cross-domain data fusion in urban computing: Taxonomy, advances, and outlook. *Information Fusion, 113*, 102606.