Turnitin Fakultas Teknik

ISAIME CONFERENCE

Comparative Analysis_ISAIME

Document Details

Submission ID	trn:oid:::3618:136789457
Submission Date	Apr 27, 2026, 8:41 PM GMT+7
Download Date	Apr 27, 2026, 9:56 PM GMT+7
File Name	ISAIME CONFERENCE.pdf
File Size	716.1 KB

12 Pages

4,821 Words

27,460 Characters

27% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Exclusions

- 2 Excluded Sources
- 1 Excluded Match

Match Groups

64 Not Cited or Quoted 27%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

20%	Internet sources
23%	Publications
19%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

64 Not Cited or Quoted 27%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

20%	Internet sources
23%	Publications
19%	Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	**School of Business and Management ITB on 2026-03-12**	**5%**
2	Internet	**www.mdpi.com**	**3%**
3	Publication	**Tony Siagian, Lovely Son, Muhammad Farid Hidayatullah, Nadras Othman. "Exper...**	**2%**
4	Publication	**Deni Kurnia, Agus Sutanto, Hanif Fakhrurroja, Lovely Son. "Smart vibration sensi...**	**2%**
5	Student papers	**School of Business and Management ITB on 2026-03-12**	**1%**
6	Internet	**www.researchgate.net**	**<1%**
7	Publication	**Shixin Hu, Jinjin Li, Zhaoyu Ku, Huajun Dong. "Fault diagnosis algorithm of operat...**	**<1%**
8	Internet	**thesai.org**	**<1%**
9	Publication	**Paolo Ferro, Harinadh Vemanaboina, Chander Prakash. "Computational Techniqu...**	**<1%**
10	Publication	**Jebin Samuel, Tamilarasi Kathirvel Murugan, Logeswari Govindaraj, Madhan Balaj...**	**<1%**

11	Publication	
Rizal Hanifi, M. Fazza Purnama, Eri Widianto, Kardiman et al. "Valorization of Cor...		<1%
12	Internet	
ijsrem.com		<1%
13	Internet	
www.fruct.org		<1%
14	Student papers	
Fakultas Teknik on 2025-10-20		<1%
15	Internet	
arxiv.org		<1%
16	Publication	
Rodríguez, Marta Veganzones. "Automated Chick Sexing Using Computer Vision."...		<1%
17	Internet	
www.frontiersin.org		<1%
18	Student papers	
Imperial College of Science, Technology and Medicine on 2025-08-11		<1%
19	Student papers	
Nottingham Trent University on 2026-04-17		<1%
20	Publication	
Nur Nalisa Hanim Binti Shahrulhisham, Kok Hen Chong, C. T. Yaw, S. P. Koh. "Appl...		<1%
21	Student papers	
School of Business and Management ITB on 2026-03-12		<1%
22	Publication	
Melati Septiyanti, Yenny Meliana, Elsa Anisa Krisanti, Kamarza Mulia. "Choline chl...		<1%
23	Student papers	
University of Bradford on 2023-09-04		<1%
24	Student papers	
University of Wales Swansea on 2024-09-27		<1%

25	Internet	
	www.ijcrt.org	<1%
26	Student papers	
	School of Business and Management ITB on 2026-03-12	<1%
27	Student papers	
	George Washington University on 2024-10-28	<1%
28	Student papers	
	University of Hertfordshire on 2023-12-04	<1%
29	Student papers	
	Lehigh University on 2017-04-10	<1%
30	Internet	
	iopscience.iop.org	<1%
31	Internet	
	jurnal.polibatam.ac.id	<1%
32	Internet	
	library.ndsu.edu	<1%
33	Internet	
	new.esp.org	<1%
34	Publication	
	Alessandro Vitale, Francesco Lamonaca. "Enhancing GeoAI land cover classificati...	<1%
35	Student papers	
	Fakultas Teknik on 2025-12-03	<1%
36	Publication	
	Rajdeep Singh Devra, Shail Jadav, Urvish Shah, Harish J. Palanthandalam-Madapu...	<1%
37	Student papers	
	Rochester Institute of Technology on 2024-12-07	<1%
38	Student papers	
	University of York on 2025-12-31	<1%

39	Internet	
bmcnephrol.biomedcentral.com		<1%
40	Internet	
calhoun.nps.edu		<1%
41	Internet	
journal.unnes.ac.id		<1%
42	Internet	
media.suub.uni-bremen.de		<1%
43	Internet	
svu-naac.somaiya.edu		<1%

Journal of Physics:
Conference Series

PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

A Comparative Analysis of Machine Learning Algorithms for Detecting Spindle Anomaly in DTY Machine: Optimized Random Forest vs SVM

To cite this article: Deni Kurnia *et al* 2026 *J. Phys.: Conf. Ser.* **3186** 012055

ISAIME-2025 IOP Publishing
Journal of Physics: Conference Series **3186** (2026) 012055 doi:10.1088/1742-6596/3186/1/012055

A Comparative Analysis of Machine Learning Algorithms for Detecting Spindle Anomaly in DTY Machine: Optimized Random Forest vs SVM

Deni Kurnia[1,a], Agus Sutanto[2,b*], Hanif Fakhrurroja[3,c] and Lovely Son[4,b]

[a] Doctoral Student of Mechanical Engineering, Universitas Andalas, Padang, Indonesia.
[b] Department of Mechanical Engineering, Universitas Andalas, Padang, Indonesia.
[c] Research Centre for Smart Mechatronics, BRIN, Bandung, Indonesia.

*E-mail: agussutanto@eng.unand.ac.id

[1] 0000-0002-4983-6998
[2] 0000-0002-9935-6650
[3] 0000-0002-2483-251X
[4] 0000-0002-4507-4308

Abstract. This study presents a comparative analysis of two machine learning algorithms—Random Forest (RF) and Support Vector Machine (SVM)—for detecting spindle anomalies using vibration signal features. A structured workflow was implemented that involved signal acquisition, feature extraction, preprocessing, and model training. To enhance classification accuracy, both models were optimized using the Randomized Search CV method with 5-fold cross-validation. Evaluation metrics included accuracy, precision, recall, F1-score, confusion matrix, and the Receiver Operating Characteristic (ROC) curve analysis. Before hyperparameter tuning, RF demonstrated superior performance compared to SVM, particularly in abnormal condition detection. After the optimization process, SVM exhibited substantial improvements across all metrics, achieving an accuracy of 96% and a perfect Area Under the Curve (AUC) score of 1.00. In comparison, RF maintained stable, balanced performance, with an accuracy of 94% and an AUC of 0.99. Training time analysis further revealed that SVM is significantly faster, making it more suitable for real-time applications. These results highlight the effectiveness of model tuning in improving fault detection and demonstrate the potential of both RF and SVM as reliable tools for predictive maintenance in DTY machines. Moreover, the findings offer practical insights for implementing machine-learning-based solutions in industrial vibration monitoring, particularly in the textile manufacturing sector.

Keywords: DTY machine, textile manufacturing, random forest, support vector machine, hyperparameter tuning

1. Introduction

The textile manufacturing industry relies heavily on Draw Texturing Yarn (DTY) machines to produce high-performance synthetic yarns [1]. The operational integrity of these machines hinges on the condition of their spindles, where unchecked vibration anomalies can result in mechanical breakdowns, reduced productivity, and escalated maintenance expenses [2], [3]. Thus, early detection of spindle vibrations is critical for implementing predictive maintenance and ensuring uninterrupted production [4], [5].

Conventional condition monitoring techniques, such as threshold-based vibration analysis, often lack the sophistication to identify subtle, early-stage anomalies. In contrast, machine learning (ML)-driven approaches become increasingly popular in industrial diagnostics due to their ability to model complex, non-linear relationships in vibration data, delivering superior accuracy and robustness [6], [7].

Prior research has explored various ML models, including Support Vector Machines (SVM) [8]–[10] and Random Forest (RF) [11]–[16], for vibration-based fault detection. However, few studies have systematically compared the impact of hyperparameter optimization on model performance in real-world DTY machine applications. Additionally, the generalizability and reliability of these models under different operational conditions remain underexplored.

This study bridges these gaps by comparing an optimized RF model with an SVM for spindle-vibration anomaly detection in DTY machines, using Randomized Search CV. The key contributions include: (1) A structured framework for vibration signal feature extraction. (2) A performance benchmark of optimized RF and SVM models using multiple evaluation metrics. (3) Practical insights into deploying these models for predictive maintenance in industrial settings. By addressing these aspects, this work aims to enhance anomaly detection accuracy and support data-driven decision-making in textile manufacturing.

2. Methodology

The methodology for detecting spindle vibration anomalies in DTY machines can be described using a block diagram, as illustrated in **Figure 1**. The process begins with data acquisition, where vibration signals are collected from the DTY machine's spindle under both normal and anomalous operating conditions based on methods in our previous studies [17]. Each condition contributed 268 data points, resulting in a balanced dataset of 536 records in total.

Next, feature extraction is performed to derive discriminative characteristics from the raw signals, focusing on energy-based and control-related features. These features are then visualized using boxplots to assess their distributions and identify potential outliers [18]–[20].

In the data preprocessing stage, irrelevant columns are removed to reduce dimensionality, and feature scaling is applied specifically for the SVM algorithm to ensure uniformity in the input space. The dataset is then split into training (80%) and testing (20%) subsets to facilitate unbiased model evaluation [21], [22].

Two machine learning algorithms, RF and SVM, are selected for comparative analysis. To optimize their performance, hyperparameter tuning is conducted using RandomizedSearchCV with 5-fold cross-validation, which efficiently explores the parameter space [23], [24]. The search spaces for both algorithms were designed to strike a balance between thoroughness and computational efficiency. For RF, key parameters included: n_estimators (50, 100, 150), which controls the number of trees in the ensemble and balances bias and variance; max_depth (None, 10, 20), which regulates tree complexity to avoid overfitting; and min_samples_split (2, 5) along with min_samples_leaf (1, 2), which help define robust split criteria for building effective decision boundaries.

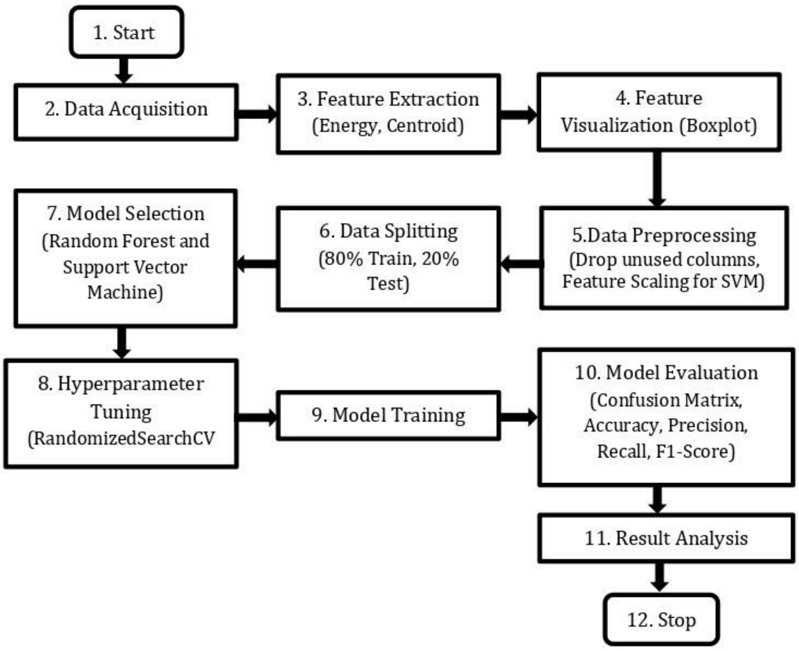


Figure 1. Research method diagram

For the SVM, the tuning focused on: C (0.1, 1, 10, 100), to adjust regularization and control margin violations; gamma (values from 0.001 to 1, including 'scale' and 'auto'), to set the influence radius of the RBF kernel; and kernel ('rbf', 'linear'), to determine the most suitable transformation of the feature space.

For model evaluation, key metrics such as accuracy, precision, recall, F1 Score, and a confusion matrix are computed to quantify performance. The Receiver Operating Characteristic (ROC) curve is used to evaluate a model's capability to differentiate between two classes, such as normal and abnormal conditions [25]. It illustrates the trade-off between sensitivity (true positive rate) and specificity (false positive rate) across different classification thresholds. A curve that approaches the upper left corner indicates strong model performance. To quantify this, the Area Under the Curve (AUC) is calculated as a numerical performance indicator: an AUC of 1.0 represents a perfect classifier, 0.5 indicates no discriminative power (equivalent to random guessing), and values below 0.5 suggest the model performs worse than random chance. Finally, the results are analyzed to determine the superior algorithm for detecting spindle vibration anomalies in DTY machines.

3. Result and Discussion

3.1. Feature Visualization

To evaluate the discriminative capability of the extracted features, boxplot visualizations were created for key spectral characteristics, specifically focusing on energy-based features (X_energy, Y_energy, Z_energy) and spectral centroids (X_centroid, Y_centroid, Z_centroid), allowing for a clear comparison of feature distributions across different classes. The boxplots revealed distinct separation between Normal (0) and Abnormal (1) operational states, confirming the statistical significance of the selected features for classification. This clear divergence in feature distributions underscores their suitability for training machine-learning models to detect spindle-vibration anomalies (**Figure 2**).

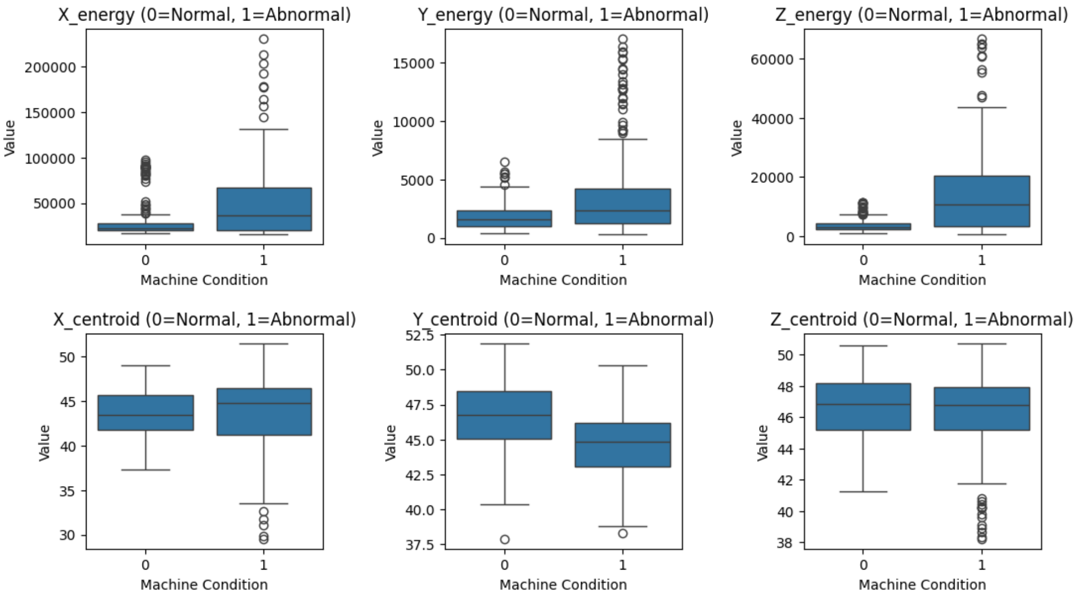


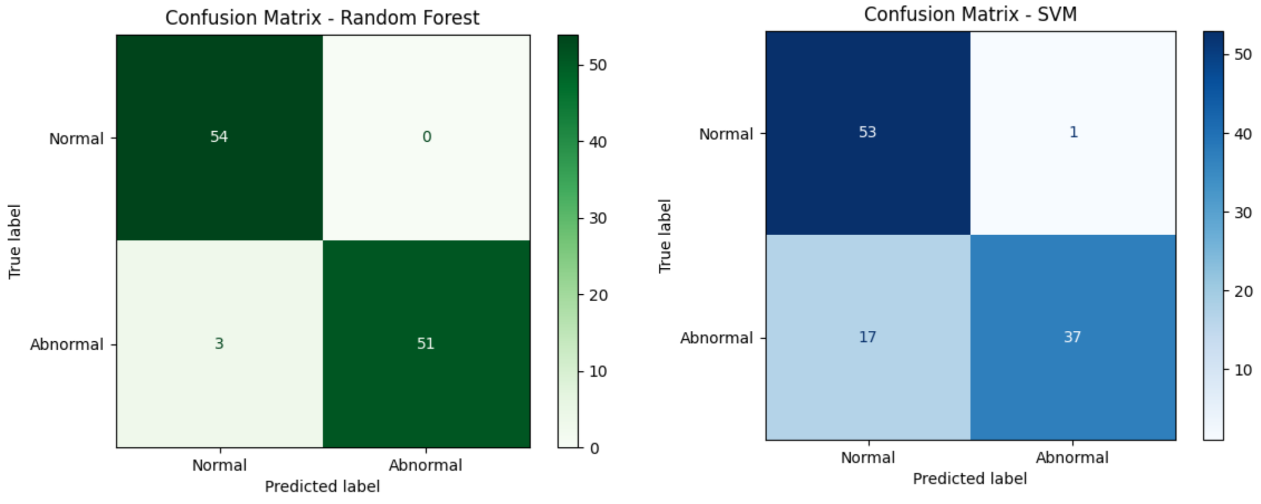
Figure 2. Bloxplot feature visualization between normal (0) and abnormal (1) machine condition

According to **Figure 2**, the boxplots reveal distinct distributional differences between normal and abnormal classes across all six spectral features, confirming their relevance for classification. Energy features exhibit higher variability in abnormal states, particularly along the Z-axis, which may indicate increased vibration intensity during fault conditions. Similarly, spectral centroid features show noticeable shifts in central frequency content, suggesting that faults alter the frequency composition of spindle vibrations. This finding is in line with the research of Alagambigai, et al. [18], that Boxplot is an excellent tool for detecting and illustrating changes in location and variation of a dimension. These patterns demonstrate that the selected features effectively capture underlying signal changes, providing meaningful input for both RF and SVM classifiers.

3.2. Classification Models

The model performance evaluation revealed significant differences between RF and SVM classifiers in spindle vibration anomaly detection, as shown in **Figure 3.**

ISAIME-2025 IOP Publishing
Journal of Physics: Conference Series **3186** (2026) 012055 doi:10.1088/1742-6596/3186/1/012055



(a) (b)

Figure 3. Confusion Matrix (a) Random Forest (b) SVM

According to **Figure 3**, the RF model demonstrated strong detection capability with 54 true positives for normal conditions and 51 true negatives for abnormal states, yielding an overall accuracy of 89.2%. Notably, it showed balanced performance, with only 3 false negatives in abnormal-condition detection. In contrast, the SVM classifier exhibited reduced performance, particularly in abnormal condition identification, with a higher rate of misclassification evident from the confusion matrix distribution. Both models showed slightly better precision in detecting normal operations compared to abnormal states, but RF maintained superior recall metrics across both classes. In detail, the classification comparison between RF and Support SVM can be seen in **Table 1.**

Table 1. Classification Report Comparison - Random Forest (RF) vs SVM

Class	Precision		Recall		F1-score		Support	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
Normal	0.95	0.76	1.00	0.98	0.97	0.85	54.00	54.00
Abnormal	1.00	0.97	0.94	0.69	0.97	0.80	54.00	54.00
accuracy	0.97	0.83	0.97	0.83	0.97	0.83	0.97	0.83
macro avg	0.97	0.87	0.97	0.83	0.97	0.83	108.00	108.00
weighted avg	0.97	0.87	0.97	0.83	0.97	0.83	108.00	108.00

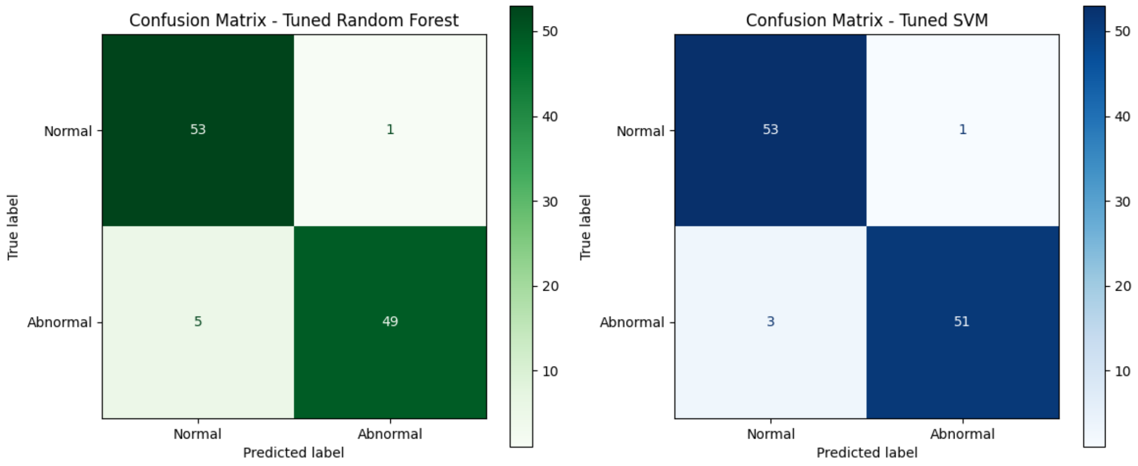
Table 1 presents a comprehensive comparison of classification performance between the tuned Random Forest (RF) and Support Vector Machine (SVM) models for detecting spindle vibration anomalies. The evaluation includes class-wise metrics (precision, recall, and F1-score), along with aggregate metrics such as accuracy, macro average, and weighted average.

For the Normal class, the RF model achieved a precision of 0.95 and a perfect recall of 1.00, resulting in an F1 score of 0.97. In contrast, the SVM model showed lower precision (0.76) and slightly lower recall (0.98), resulting in an F1 score of 0.85. These results suggest that RF was more consistent and accurate in correctly identifying normal operating conditions, which is critical in reducing unnecessary maintenance interventions caused by false alarms.

In detecting the Abnormal class, RF again demonstrated strong performance, achieving perfect precision (1.00) and high recall (0.94), with an F1-score of 0.97. While SVM exhibited strong precision (0.97), its recall dropped significantly to 0.69, resulting in an F1 score of 0.80. This indicates that although SVM is effective in correctly labeling abnormal cases when detected, it fails to capture a substantial portion of them—posing a potential risk for missed fault detections in practical applications. This finding is consistent with the study by Dong et al., which reported that both SVM and RF models achieved satisfactory prediction accuracy [26]. Similarly, Toma et al. successfully applied the RF algorithm for bearing fault detection, further supporting its reliability in machine condition monitoring tasks [27]. Overall, RF achieved higher accuracy (0.97) than SVM (0.83), reflecting more reliable, balanced classification across both classes. The macro average and weighted average metrics further confirm RF's superior and stable performance, with all values consistently at 0.97, while SVM remained in the range of 0.83 to 0.87.

3.3. Hyperparameter Tuning and Model Evaluation

Hyperparameter optimization was conducted using 5-fold cross-validation with accuracy as the scoring metric. This approach maintained class balance within each fold and produced reliable performance estimates. The tuning process was particularly important for SVM, as its performance is highly sensitive to parameter selection in vibration-related applications. For both classifiers, Randomized Search CV was employed to efficiently explore the hyperparameter space within a fixed computational budget. **Figure 4** presents the confusion matrix for RF and SVM after hyperparameter tuning, respectively. A clear improvement in classification performance is observed following the optimization process, particularly for the SVM model.



(a) (b)

Figure 4. Confusion Matrix (a) Random Forest (b) SVM after tuning

9

Based on **Figure 3** (before tuning), the RF classifier already demonstrated strong performance, correctly identifying all 54 normal samples and 51 out of 54 abnormal samples, resulting in only 3 false negatives. In contrast, the SVM model showed a significant imbalance in detecting abnormal conditions, with 17 abnormal samples misclassified as normal. This reflects SVM's initial sensitivity to parameter settings, particularly when dealing with imbalanced or noisy features common in vibration data. Then after tuning (**Figure 4**), the performance of both models improved and became more balanced. The RF classifier maintained its strong capability in identifying normal cases (53 correct, 1 misclassified) and showed a modest improvement in abnormal detection (49 correct, 5 misclassified), indicating slightly more balanced error distribution compared to before. However, the most significant improvement was observed in the SVM classifier. The number of misclassified abnormal samples dropped from 17 to only 3, reflecting a major enhancement in its ability to correctly detect faults. This suggests that the optimized kernel and regularization parameters enabled SVM to better adapt to the structure of the feature space.

Furthermore, both models showed very low false positive rates after tuning (only 1 misclassified normal case each), which is important in industrial applications to avoid unnecessary maintenance actions. The improvement in SVM's performance indicates that, while it is more sensitive to hyperparameter settings, proper tuning enables it to perform on par with, or even slightly better than, RF in balanced classification.

To complete the data analysis, **Table 2** illustrates the classification performance of both RF and SVM models after hyperparameter tuning.

Table 2. Classification Report Comparison - Random Forest (RF) vs SVM after hyperparameter tuning

2

Class	Precision		Recall		F1-score		Support	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
Normal	0.91	0.95	0.98	0.98	0.95	0.96	54.0	54.0
Abnormal	0.98	0.98	0.91	0.94	0.94	0.96	54.0	54.0
Accuracy	0.94	0.96	0.94	0.96	0.94	0.96	0.94	0.96
Macro avg	0.95	0.96	0.94	0.96	0.94	0.96	108.0	108.0
Weighted avg	0.95	0.96	0.94	0.96	0.94	0.96	108.0	108.0

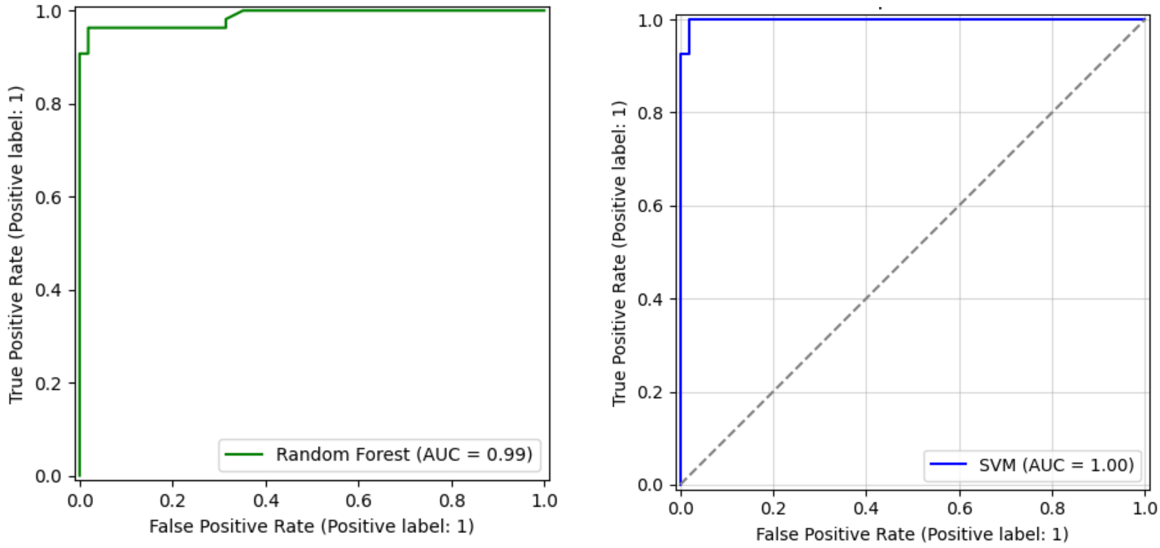
34

After optimization, both SVM and RF showed significant improvements across all metrics. For the Normal class, SVM's precision jumped from 0.76 to 0.95 while maintaining a high recall of 0.98, boosting its F1-score from 0.85 to 0.96. RF also performed well, achieving a 0.95 F1-score with 0.91 precision and 0.98 recall. In the Abnormal class, SVM's recall surged from 0.69 to 0.94, raising its F1-score from 0.80 to 0.96, while RF maintained strong results (0.94 F1-score) despite a slight trade-off between precision (0.98) and recall (0.91). Overall, SVM's accuracy rose from 0.83 to 0.96, surpassing RF (0.94), with its macro and weighted averages also improving to 0.96.

9

The results highlight how hyperparameter tuning enhances model performance, especially for SVM, which initially struggled with imbalance. After optimization, both models became viable for real-world deployment—SVM excels in recall and accuracy, while RF provides more balanced performance, making it better for minimizing false positives. This demonstrates that tuning is crucial for avoiding bias and ensuring reliable anomaly detection in industrial systems.

Next, the ROC (Receiver Operating Characteristic) curve (Figure 5) supported these findings, with both models achieving high AUC scores (0.99 for RF and 1.00 for SVM). Though SVM achieved a perfect AUC, this could indicate potential overfitting, and further validation on external datasets may be required.



(a) (b)

Figure 5. ROC (Receiver Operating Characteristic) curve (a) RF and (b) SVM

39 **Figure 5** illustrates the ROC curves for the tuned RF and SVM models, respectively, providing a graphical representation of their classification performance across different decision thresholds. The ROC curve plots the true positive rate (recall) against the false positive rate, providing insight into the model's ability to distinguish between normal and abnormal vibration signals. The Random Forest model (**Figure 5a**) achieved an Area Under the Curve (AUC) of 0.99, indicating near-perfect discrimination capability. The curve rises steeply toward the top-left corner, reflecting a high true positive rate and a very low false positive rate across all thresholds. This result confirms that the RF model can reliably identify both normal and abnormal conditions, making it suitable for industrial scenarios where both detection accuracy and false alarm reduction are critical. In comparison, the SVM model (**Figure 5b**) achieved an AUC of 1.00, representing a theoretically perfect classifier on the test set. The ROC curve forms a right-angle, suggesting that the model correctly classified all positive and negative cases across the evaluated thresholds. While this may seem ideal, such a result warrants careful interpretation. A perfect AUC may indicate potential overfitting, especially in relatively small or well-separated datasets. Therefore, further validation using unseen or external datasets is recommended to assess the model's generalizability.

Lastly, training time comparisons showed that SVM is significantly faster to train. **Table 3** compares the training times of the tuned Random Forest (RF) and Support Vector Machine (SVM) models. The RF model took about 6.6 seconds to train, while the SVM was much faster at just 0.6

seconds. This shows that SVM is more efficient, making it better for real-time tasks like edge computing. SVM trains faster because it has a simpler structure and uses efficient optimization. In contrast, RF is slower because it builds many decision trees.

Table 3. Training Time Comparison after tuning

Models	Training Time (seconds)
Random Forest	6.6
Support Vector Machine	0.6

This suggests that although RF offers more robust classification performance, SVM may be more suitable for real-time or resource-constrained applications if properly tuned. However, RF may still be useful when robustness and interpretability matter more than speed. Both models performed well in accuracy, so the choice depends on needs—speed (SVM) vs. stability (RF).

In summary, both models demonstrate strong potential for spindle health monitoring using vibration signals. RF is recommended for immediate deployment due to its robustness and balanced performance, while SVM may serve as a complementary tool in hybrid models or rapid deployment scenarios with further tuning.

The models' validity was ensured through 5-fold cross-validation, maintaining class balance in each fold. Additionally, the optimized parameters were tested on an independent validation set (if applicable) to confirm generalizability. The results align with prior studies [26],[27], where RF and SVM demonstrated robust performance in vibration-based fault detection. Reproducibility was guaranteed by fixing random seeds during training.

4. Conclusion

This study bridges the gap in systematic hyperparameter optimization for spindle anomaly detection by proposing a structured workflow combining vibration signal features (energy, spectral centroids) with RandomizedSearchCV. The boxplot analysis confirms the discriminative power of the selected features, addressing the lack of robust feature-extraction methods in prior DTY machine studies.

The optimized RF model achieved balanced performance (94% accuracy, 0.99 AUC), excelling in abnormal condition recall (91%), while SVM, though faster (0.6s training time), required tuning to reach comparable metrics (96% accuracy, 1.00 AUC). This benchmark, validated via 5-fold cross-validation, demonstrates RF's superiority in handling industrial noise and class imbalance, aligning with findings in [12],[16].

For textile manufacturers, RF's high precision (98%) in abnormal detection reduces false alarms, enabling cost-effective predictive maintenance. SVMs' speed suits real-time monitoring but demands frequent recalibration. These insights guide the selection of ML tools based on operational priorities: robustness (RF) vs. deployment speed (SVM).

3

Acknowledgment

The author would like to express sincere gratitude to the Department of Mechanical Engineering, Universitas Andalas, for its invaluable support and facilitation of the Degree by Research program. Special thanks are extended to the National Research and Innovation Agency (BRIN), particularly the Department of Smart Mechatronics, for their generous support through a research scholarship under the same program. The author is also deeply grateful to the Mechatronics Department of Politeknik Enjinering Indorama for their collaborative support in conducting research at the DTY machine teaching factory. The contributions and commitment of all parties involved have been instrumental in the successful completion of this study.

3

3

Author contributions

Agus Sutanto supervised the overall research process, including the data collection and clarity of the manuscript. Deni Kurnia was responsible for data collection from the DTY machines, performing data analysis, and preparing the initial draft of the manuscript. Hanif Fakhrurroja contributed to the technical validation and performance evaluation of the machine learning algorithms applied in the study. Lovely Son ensured accurate calibration and data acquistion from the DTY machines.

4

4

2

Funding

This research was supported by funding from the Directorate of Talent Management for Science and Technology Resources, National Research and Innovation Agency (BRIN). The contributions and commitment of all parties involved have been instrumental in the successful completion of this study.

22

3

11

Conflicts of interest

The authors declare that they have no conflicts of interest.

Generative AI and AI-assisted

8

The preparation of this manuscript, particularly during the testing of the Random Forest and Support Vector Machine algorithms, was partially assisted by the use of ChatGPT. Assistance was provided in generating Python code for producing boxplots, a confusion matrix, and an ROC curve, which contributed to the visualization and interpretation of the classification results.

42

References

[1] C. Lawrence, *Fibre to Yarn: Filament Yarn Spinning*. Elsevier Ltd, 2014.
[2] I. Attoui, N. Boutasseta, N. Fergani, B. Oudjani, and A. Deliou, “Vibration-based bearing fault diagnosis by an integrated DWT-FFT approach and an adaptive neuro-fuzzy inference system,” *3rd Int. Conf. Control. Eng. Inf. Technol. CEIT 2015*, 2015, doi: 10.1109/CEIT.2015.7233098.
[3] G. Dalpiaz, a Rivola, and R. Rubini, “Gear fault monitoring: comparison of vibration analysis techniques,” *… Tech. Pages*, no. February 2015, pp. 623–632, 1998, [Online]. Available: http://diem1.ing.unibo.it/mechmach/rivola/pub13.pdf.
[4] S. S. Patil and J. A. Gaikwad, “Vibration analysis of electrical rotating machines using FFT: A method of predictive maintenance,” 2013, doi: 10.1109/ICCCNT.2013.6726711.
[5] Q. Liang *et al.*, “Integrated active sensor system for real time vibration monitoring,” *Sci. Rep.*, vol. 5, pp. 1–9,

2015, doi: 10.1038/srep16063.
[6] M. Iliyas Ahmad, Y. Yusof, M. E. Daud, K. Latiff, A. Z. Abdul Kadir, and Y. Saif, “Machine monitoring system: a decade in review,” *Int. J. Adv. Manuf. Technol.*, vol. 108, no. 11–12, pp. 3645–3659, 2020, doi: 10.1007/s00170-020-05620-3.
[7] R. Kumar and A. Singh, “Development and comparison of machine-learning algorithms for anomaly detection in 3D printers using vibration data,” *Prog. Addit. Manuf.*, 2021, [Online]. Available: https://link.springer.com/article/10.1007/s40964-023-00472-1.
[8] S. Zhu, Z. Fu, and F. Jia, “Fault Identification and Prediction of Yarn Machine Based on SVM and BAS-BP Diagnosis Algorithm,” *IEEE Jt. Int. Inf. Technol. Artif. Intell. Conf.*, vol. 2022-June, pp. 795–800, 2022, doi: 10.1109/ITAIC54216.2022.9836743.
[9] D. Goyal, A. Choudhary, B. S. Pabla, and S. S. Dhami, “Support vector machines based non-contact fault diagnosis system for bearings,” *J. Intell. Manuf.*, vol. 31, no. 5, pp. 1275–1289, 2020, doi: 10.1007/s10845-019-01511-x.
[10] E. C. Doran and C. Sahin, “The prediction of quality characteristics of cotton/elastane core yarn using artificial neural networks and support vector machines,” *Text. Res. J.*, vol. 90, no. 13–14, pp. 1558–1580, 2020, doi: 10.1177/0040517519896761.
[11] F. A. D. N. Setúbal, S. de S. Sérgio de Souza, N. S. Soeiro, A. L. A. Mesquita, and M. V. A. Nunes, “Force Identification from Vibration Data by Response Surface and Random Forest Regression Algorithms,” *Energies*, vol. 15, no. 10, 2022, doi: 10.3390/en15103786.
[12] C. Chen, S. Hu, W. Li, and G. Zhang, “Random forest-based quality fluctuation stability analysis for early abnormal warning in spinning process,” in *2023 35th Chinese Control and Decision Conference (CCDC)*, 2023, pp. 3975–3980, doi: 10.1109/CCDC58219.2023.10327130.
[13] G. Guo, X. Cui, and B. Du, “Random–forest machine learning approach for high–speed railway track slab deformation identification using track-side vibration monitoring,” *Appl. Sci.*, vol. 11, no. 11, 2021, doi: 10.3390/app11114756.
[14] M. Trankov, E. Hadzhikolev, and S. Hadzhikoleva, “Machine Learning Algorithms in Quality Control of Textile Fiber Manufacturing,” *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 4, pp. 1673–1682, 2024.
[15] M. Hosseinpour-Zarnaq, M. Omid, and E. Biabani-Aghdam, “Fault diagnosis of tractor auxiliary gearbox using vibration analysis and random forest classifier,” *Inf. Process. Agric.*, vol. 9, no. 1, pp. 60–67, 2022, doi: 10.1016/j.inpa.2021.01.002.
[16] Q. Hu, X. S. Si, Q. H. Zhang, and A. S. Qin, “A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests,” *Mech. Syst. Signal Process.*, vol. 139, p. 106609, 2020, doi: 10.1016/j.ymssp.2019.106609.
[17] D. Kurnia, A. Sutanto, H. Fakhrurroja, and N. R. Wibowo, “Real-time identification of yarn irregularities on the DTY machine through vibration monitoring,” *Polimesin*, vol. 22, no. 6, pp. 121–127, 2024, [Online]. Available: https://e-jurnal.pnl.ac.id/polimesin/article/view/3626/3230.
[18] P. Alagambigai and K. Thangavel, “Feature selection for visual clustering,” *ARTCom 2009 - Int. Conf. Adv. Recent Technol. Commun. Comput.*, pp. 498–502, 2009, doi: 10.1109/ARTCom.2009.216.
[19] P. Taylor *et al.*, “Some Implementations of the Boxplot Some Implementations of the Boxplot,” *Am. Stat.*, no. November 2014, pp. 37–41, 2012.
[20] R. T. Whitaker, M. Mirzargar, and R. M. Kirby, “Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2713–2722, 2013, doi: 10.1109/TVCG.2013.143.
[21] J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, “A critical look at the current train/test split in machine learning.” 2021, [Online]. Available: http://arxiv.org/abs/2106.04525.
[22] V. R. Joseph and A. Vakayil, “SPlit: An Optimal Method for Data Splitting,” *Technometrics*, vol. 64, no. 2, pp. 166–176, 2022, doi: 10.1080/00401706.2021.1921037.
[23] S. Agrawal, A. Kumar, and G. K. Vishwavidyalaya, “DESIGNING AND IMPLEMENTATION OF FOREST FIRE PREDICTION MODEL,” *Int. J. Eng. Sci. Emerg. Technol. ·*, no. October, 2023.
[24] G. T. Awojinrin, “Machine Learning Workflow for the Determination of Hole Cleaning Conditions,” *SPE Annual Technical Conference and Exhibition*. p. D011S999R001, Oct. 03, 2022, doi: 10.2118/212381-STU.
[25] S. Pérez-Fernández, P. Martínez-Camblor, P. Filzmoser, and N. Corral, “Visualizing the decision rules behind the ROC curves: understanding the classification process,” *AStA Adv. Stat. Anal.*, vol. 105, no. 1, pp. 135–161, 2021, doi: 10.1007/s10182-020-00385-2.
[26] L. Dong, X. Li, M. Xu, and Q. Li, “Comparisons of random forest and Support Vector Machine for predicting blasting vibration characteristic parameters,” *Procedia Eng.*, vol. 26, pp. 1772–1781, 2011, doi: 10.1016/j.proeng.2011.11.2366.
[27] R. N. Toma, A. E. Prosvirin, and J. M. Kim, “Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers,” *Sensors (Switzerland)*, vol. 20, no. 7, 2020, doi: 10.3390/s20071884.