Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1142

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Balinese Rindik Note Recognition Using Discrete Wavelet Transform and Support Vector Machine

I Made Jepri Antara[1], Ni Gusti Ayu Putu Harry Saptarini[1*✉], Putu Desiana Wulaning Ayu[1], I Putu Astya Prayudha[1]
[1]Information Technology Department, Politeknik Negeri Bali, Badung, Indonesia
**Corresponding Author: ayu.harry@pnb.ac.id*

Article Information

Article history:

No. 1142
Rec. May 01, 2026
Rev. November 06, 2026
Acc. November 23, 2026
Pub. November 26, 2026
Page. 151–175

Keywords:

- *Rindik*
- *Discrete Wavelet Transform*
- *RMS Normalization*
- *Support Vector Machine*
- *Machine Learning*

ABSTRACT

Automatic recognition of Balinese Rindik notes is challenging because bamboo resonance signals are non stationary and striking intensity varies from stroke to stroke. This study investigates a machine learning pipeline that combines the Discrete Wavelet Transform (DWT), Root Mean Square (RMS) normalization applied to the wavelet sub bands, and a Support Vector Machine (SVM). After silence trimming and amplitude normalization, level 3 Daubechies 1 (db1) decomposition is applied, and four statistical features (mean, standard deviation, energy, and entropy) are extracted from each of the four sub bands (A3, D3, D2, D1), yielding a 16 dimensional feature vector. An SVM with a Radial Basis Function (RBF) kernel classifies the five Panca Nada notes (Ding, Dong, Deng, Dung, and Dang). Of 500 self recorded monophonic samples, 400 were used for training and 100 were held out for testing. Five fold stratified cross validation on the training samples gave a mean accuracy of 99.25% ± 0.61% this is a training phase result, not the final test result. Evaluated once on the 100 held out samples, the final model achieved 100.00% accuracy, precision, recall, and F1 score (ROC AUC = 1.0000), and file hashing confirmed that no recording was shared between the two subsets. The ablation study shows that DWT significantly improves accuracy over time domain features ($p = 0.002$), whereas the additional effect of sub band RMS normalization is not statistically significant on this controlled dataset ($p = 0.374$). An FFT baseline reached 85.00% test accuracy, while an MFCC baseline matched the 100.00% of the proposed method. Because the data come from a single performer under low noise conditions, these results should not be generalized to uncontrolled settings. The system is deployed as a Streamlit web application to support the digital preservation of traditional Balinese instruments.

How to Cite:

Antara, I. M. J., & et al. (2026). Balinese Rindik Note Recognition Using Discrete Wavelet Transform and Support Vector Machine. Jurnal Teknologi Informasi Dan Pendidikan, 20(1), 151–175. https://doi.org/10.24036/jtip.v20i1.1142

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1142

This open-access article is distributed under the Creative Commons Attribution-ShareAlike 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2023 by Jurnal Teknologi Informasi dan Pendidikan.

1. INTRODUCTION

Indonesia is home to various traditional arts, including musical instruments that reflect local cultural identities. Balinese gamelan, specifically the bamboo based Rindik instrument, has unique acoustic characteristics and plays a key role in Balinese culture. The instrument produces complex acoustic signals consisting of fundamental frequency components and harmonics that form a distinctive timbre [1].

Recent technological advancements have helped digitize cultural preservation through intelligent systems and multimedia platforms [17]. However, automatic recognition of musical notes for digital notation remains challenging. This process relies heavily on human hearing, while conventional digital tuners frequently struggle to separate fundamental frequencies from complex bamboo harmonic frequencies.

Previous research indicated that fundamental frequency analysis using the Fast Fourier Transform (FFT) is important for understanding the characteristics of Rindik note blades during sound synthesis [1]. Although effective in frequency detection, FFT alone is often insufficient for accurate note recognition. It struggles when handling non stationary signals. For this reason, wavelet analysis is often preferred, because it captures time and frequency information at the same time, which has proved useful for automatically transcribing Rindik notes [2], [11]. In addition, machine learning algorithms like the Support Vector Machine (SVM) show strong performance in classifying complex musical patterns, including Balinese gamelan [3], [10]. Hybrid approaches combining Discrete Wavelet Transform (DWT) feature extraction and SVM classification have also been successfully applied in audio detection systems [4], [12]. More broadly, wavelet domain features have been used in content based music classification [19], [20], and multiresolution wavelet analysis is a well established tool for non stationary signals [14]. Automatic music transcription itself remains an open research problem [21].

Even with these advancements, existing studies still have difficulty with amplitude variability. This instability is mainly caused by different striking forces on the bamboo blades, which reduces feature stability and may affect classification accuracy. Methods such as Mel Frequency Cepstral Coefficients (MFCC) [16] have proven effective in many audio classification tasks [5], [6], but they were originally developed for speech and do not directly address amplitude fluctuations caused by changes in striking force. Similarly, raw Discrete Wavelet Transform (DWT) features, while capable of capturing time frequency information,

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

may remain sensitive to energy variation across recordings when applied without any normalization step [2], [11]. In the Rindik and Balinese gamelan studies reviewed here [1]–[3], [10], the emphasis has been on frequency analysis, transcription, or classifier comparison, and the effect of striking force variation on the feature vectors has not been examined explicitly.

Amplitude and energy normalization are well established ideas in audio processing, so this study does not claim that normalization itself is new. To the best of our knowledge, and within the scope of the literature reviewed, applying RMS normalization separately to each DWT sub band of bamboo percussive notes, and testing it under a leakage controlled protocol, has not been reported before. The contributions of this study are therefore fourfold (i) a DWT–RMS–SVM pipeline that represents each Rindik note as a 16 dimensional feature vector (ii) an evaluation protocol that strictly separates 5 fold stratified cross validation on 400 training samples from a single final evaluation on 100 held out test samples, supported by file level duplicate verification (iii) an ablation study with a paired t test and FFT and MFCC baselines, which reports transparently where sub band RMS normalization does and does not make a measurable difference and (iv) a Streamlit application that makes the trained model usable for digital preservation. System performance is measured using Accuracy, Precision, Recall, F1 Score, and the Receiver Operating Characteristic Area Under the Curve (ROC AUC).

Rindik is a traditional Balinese musical instrument constructed from bamboo. Technically, the sound produced by Rindik has complex wave characteristics. The instrument produces audio signals consisting of basically frequency components and harmonic structures that form its distinctive timbre [1]. Similar characteristics are also seen in other traditional musical instruments, where early transients and long resonance play a key role in sound formation [6]. As a result, traditional bamboo music audio signals need stable preprocessing stages to improve signal clarity before further analysis.

Feature extraction techniques are essential for improving classification performance in audio processing. While approaches such as Mel Frequency Cepstral Coefficients (MFCC) have been widely used in audio classification tasks [5], [6], wavelet features are often preferred for signals with strong temporal variations. The Wavelet Transform is a signal analysis method that decomposes signals into multiple components with varying frequency scales and time resolutions. Unlike the Fourier transform, which only operates in the frequency domain, the Wavelet Transform retains time information, making it highly effective for analyzing non stationary signals with time frequency localization, such as Rindik music [2]. Several studies have demonstrated the effectiveness of the Discrete Wavelet Transform (DWT) in extracting meaningful features from various audio signals, including applications in biomedical signals and mechanical system monitoring [11], [13].

To ensure the quality of the extracted features, normalization techniques play a critical role in stabilizing feature values. Root Mean Square (RMS) normalization helps reduce the impact of amplitude variations and improves the consistency of extracted

features, especially in signals affected by physical interactions such as musical instruments. In this study, DWT decomposition is complemented by RMS normalization applied to each sub band. This normalization is designed to standardize the magnitude of energy generated from varying striking forces, which should allow statistical features (such as mean and energy) to provide more stable frequency representations. Whether this design translates into a measurable accuracy gain is tested empirically in the ablation study (Section 3.4).

For classification, the Support Vector Machine (SVM) is a supervised learning algorithm that handles high dimensional and nonlinear data well [3], [15]. It operates by finding the most effective hyperplane to separate data into different classes. Previous research shows that SVM performs well when classifying Balinese gamelan genres and other audio signals because of its good data generalization [3], [8], [10]. Combining wavelet feature extraction and SVM classification is effective in recognizing sound signal patterns [4]. Other classification methods, such as Decision Tree and Hidden Markov Model (HMM), have also been explored in audio recognition tasks [7], [9] however, they were not evaluated here, because SVM is well suited to small datasets with compact feature vectors. Therefore, to handle the nonlinear Rindik audio data, the Radial Basis Function (RBF) Kernel is utilized, as it forms flexible decision boundaries to fit the frequency resonances.

2. RESEARCH METHOD

This study adopts a quantitative experimental approach to develop a machine learning system for Rindik note recognition. The system architecture follows a structured computational pipeline consisting of data acquisition, preprocessing, feature extraction, classification, and evaluation. To avoid any ambiguity between the 80/20 hold out scheme and cross validation, the evaluation is organized into two strictly separated stages 5 fold stratified cross validation, which is applied only to the training subset for learning and hyperparameter tuning, and a single final evaluation on a held out test subset that is never used during training or tuning (Section 2.6).

2.1. System Architecture Design

The Rindik note recognition software was developed using a Web Interface model. This setup allows users to communicate directly with the Audio Processing module. The system logic was documented using a system architecture diagram and a processing flowchart, and the software was built as a web application using Streamlit for real time interaction and prediction [18]. Figure 1 shows the main software components, from the web interface to the SVM classification module, ensuring a smooth data flow from user input to the final prediction. Figure 2 details the processing stages, starting from audio acquisition to confidence score validation, explaining how the program logic handles each input sequentially.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

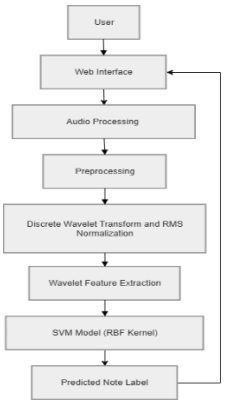


Figure 1. System architecture of the note recognition software

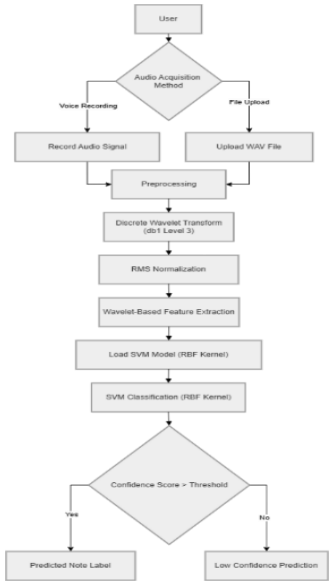


Figure 2. Flowchart of the system logic and workflow

2.2. Data Acquisition Procedure and Environment

The data acquisition process was conducted in a controlled environment with low ambient noise interference. Recordings were performed consistently at a distance of ±30 cm from the bamboo blade using a single strike technique. Unlike studies that rely on publicly available audio repositories, the dataset used in this research was self recorded by the author using a real Rindik instrument and a condenser microphone. This approach was deliberately chosen to ensure that the audio samples faithfully captured the genuine acoustic characteristics of the instrument, including its natural resonance, tonal nuances, and physical striking dynamics, qualities that may not be adequately represented in generic or synthetic audio collections. Recordings were captured through the recording module of

the web application (Tab 1, Figure 3), in which the operator selects the note label, the number of recordings per session, and the delay between consecutive recordings (2.00 s in the example shown) every file is saved automatically to the data folder of the selected note. This process yielded 500 monophonic audio recordings in .wav format with a 2 second duration and a 44,100 Hz sampling frequency, representing the five basic Panca Nada notes (Ding, Dong, Deng, Dung, and Dang) with 100 recordings per note. The partitioning into training and test subsets was carried out by the author, not predefined by any external source, using a stratified 80/20 hold out scheme 80% of the recordings were allocated for training (400 samples, 80 per note) and 20% for testing (100 samples, 20 per note). The two subsets are stored in physically separate folders, and the test subset was not accessed during cross validation, hyperparameter tuning, or model selection (Section 2.6). The size of this dataset was determined based on the number of available Panca Nada note classes and the practical constraints of a single instrument monophonic recording setup. Limitations of this dataset size are discussed further in Section 3.7.

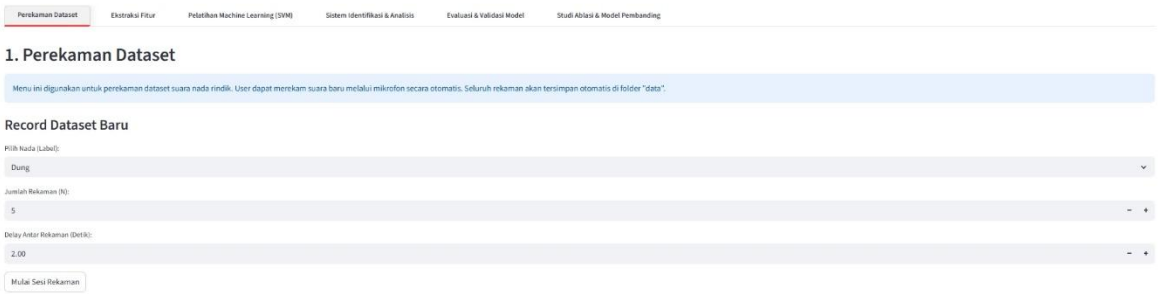


Figure 3. Recording tab of the web application

2.3. Data Preprocessing and Feature Extraction

Raw audio signals first underwent silence trimming and initial amplitude normalization to improve signal clarity. The cleaned signals were then decomposed using DWT with the Daubechies 1 (db1) mother wavelet at level 3, producing four sub bands A3 (approximation), D3, D2, and D1 (detail coefficients). DWT has been widely validated across signal analysis tasks [2], [11], [14]. The distinctive design choice examined in this study is applying RMS normalization to each wavelet sub band, which standardizes the energy magnitude before feature extraction. The RMS of a sub band with N coefficients is defined as:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad (1)$$

From each normalized sub band, four statistical features are extracted mean (μ), standard deviation (σ), energy (E), and Shannon entropy (H) across the four sub bands (A3,

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

D3, D2, D1), yielding a 16 dimensional feature vector per sample. The application saves the features of the 400 training recordings as a 400 × 16 matrix (Figure 12), and the 100 held out recordings contribute a further 100 × 16 matrix, which gives a 500 × 16 dataset overall. The extraction step also includes a switch that enables or disables RMS normalization. Their formulas are: The Mean (μ) calculates the central tendency of the frequency sub band:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

The Standard Deviation (σ) measures the dispersion of the amplitude from the mean:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (3)$$

The Energy (E) quantifies the signal strength within the specific wavelet decomposition level:

$$E = \sum_{i=1}^{N}|x_i|^2 \quad (4)$$

The Shannon Entropy (H) measures the uncertainty or complexity of the signal's energy distribution:

$$H = -\sum_{i=1}^{N} p_i \log(p_i) \quad (5)$$

Where μ is the mean value, σ is the standard deviation, and E is the signal energy. For the entropy calculation, p_i represents the probability of the i-th coefficient's energy within the distribution. These statistical features are commonly utilized in signal processing and machine learning applications [13].

2.4. Pseudocode: DWT and RMS Feature Extraction Pipeline

The complete sequence of this data processing stage is structured in Table 1. The pipeline takes the raw audio signal as the input and applies the wavelet decomposition followed by the energy normalization. Finally, the system extracts the statistical values from each sub band to form the 16 dimensional feature vector for the classification phase.

Table 1. DWT and RMS Feature Extraction Algorithm

Parameter / Step	Description / Process
Input	Audio signal *s*, wavelet 'db1', level *L*=3
Output	16 dimensional feature vector F
Step 1	Trim silence and normalize amplitude of s
Step 2	Apply DWT: [A3, D3, D2, D1] ← dwt(s, 'db1', level=3)
Step 3	For each sub band C_k in {A3, D3, D2, D1}: a. $rms_k \leftarrow \sqrt{mean(C_k^2)}$

	b. $C_{k_norm} \leftarrow \frac{c_k}{rms_k}$
	c. Extract [mean, std, energy, entropy] from C_{k_norm}
Step 4	Concatenate all 4 × 4 = 16 features → F
Step 5	Return F

2.5. SVM Classification and Evaluation

The 16 dimensional feature vectors were fed into an SVM classifier using the RBF kernel, which maps nonlinear features into higher dimensional space [3], [8], [15]. The RBF kernel function is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right) \quad (6)$$

Where $\left|\left|x_i - x_j\right|\right|^2$ represents the squared Euclidean distance between two feature vectors, and (γ) defines the kernel's influence radius. Optimal hyperparameters were identified via GridSearchCV across C ∈ {0.1, 1, 10, 100, 1000} and γ ∈ {'scale', 'auto', 0.01, 0.1, 1}. Table 2 lists the value ranges for the C and gamma parameters tested during this search process to find the most effective configuration.

Table 2. GridSearchCV Hyperparameter Search Space for SVM

Hyperparameter	Tested Values (Search Space)	Description
C (Regularization)	0.1, 1, 10, 100, 1000	Controls the penalty for misclassified data points.
Gamma (Kernel Coefficient)	'scale', 'auto', 0.01, 0.1, 1	Defines the influence radius of a single training example.
Kernel	'rbf'	The mathematical function used to map data.

The grid search was run on the 400 training samples only, using the same stratified 5 fold splits as the cross validation described below, so that no test sample could influence the choice of C and γ.

We used 5 fold stratified cross validation, applied exclusively to the 400 sample training subset, to monitor learning stability and to select hyperparameters. Stratification keeps the five note classes equally represented in every fold (16 samples per class in each validation fold). The selection of 5 folds, rather than a smaller or larger value of k, was informed by both theoretical reasoning and established practices in the field. 5 fold cross validation strikes a practical balance between bias and variance in performance estimation each fold reserves 20% of the training data for validation while training on the remaining 80%, which mirrors the widely adopted 80/20 partitioning strategy in machine learning. Beyond this general principle, several prior studies in audio classification and musical instrument recognition tasks have found that 5 fold cross validation yields stable and trustworthy evaluation outcomes without incurring excessive computational overhead [3],

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

[4], [10]. Cross validation scores are therefore reported as training phase indicators only the final performance of the system is measured on the separate 100 sample test subset (Section 2.6). The final model was evaluated using a Confusion Matrix. Table 3 outlines the standard structure of this matrix, tabulating the predicted classes against the actual classes. This layout defines the four basic parameters used for further calculations:

Table 3. Standard structure of a confusion matrix

Actual \ Predicted	Positif (Predicted Positive)	Negatif (Predicted Negative)
Positif (Actual Positive)	True Positive (TP)	False Negative (FN)
Negatif (Actual Negative)	False Positive (FP)	True Negative (TN)

Based on these parameters, the final model performance on the held out test subset was assessed using the confusion matrix and the metrics Accuracy (Eq. 7), Precision (Eq. 8), Recall (Eq. 9), and F1 Score (Eq. 10), complemented by one vs rest ROC AUC. Because the task has five classes, Precision, Recall, and F1 Score are reported per note and as macro and weighted averages.
Accuracy measures the overall correctness of the model in classifying all note classes:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Precision calculates the ratio of correctly predicted positive observations to the total predicted positives, ensuring that false notes are not misclassified as the target note:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Recall (or Sensitivity) measures the model's ability to correctly identify all actual positive instances of a specific bamboo note:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

F1 Score provides the harmonic mean of Precision and Recall, which is important for evaluating the model's performance balance, especially in multi class audio recognition:

$$F1 - Score = \frac{Precision\ x\ Recall}{Precision+Recall} \quad (10)$$

2.6. Evaluation Protocol and Data Leakage Prevention

The evaluation follows the two stage protocol shown in Figure 4. In Stage 1, the 400 training samples are used for 5 fold stratified cross validation of the pipeline (StandardScaler followed by an SVM with RBF kernel) and for GridSearchCV hyperparameter tuning. In Stage 2, the final pipeline is retrained on all 400 training samples and evaluated exactly once on the 100 held out test samples. The 5 fold cross validation and

the 80/20 hold out scheme are therefore not competing evaluation methods but complementary stages with different purposes the former estimates learning stability and supports model selection, whereas the latter provides the final performance estimate that is not biased by model selection [24]. Table 4 summarizes the class distribution of the two subsets.

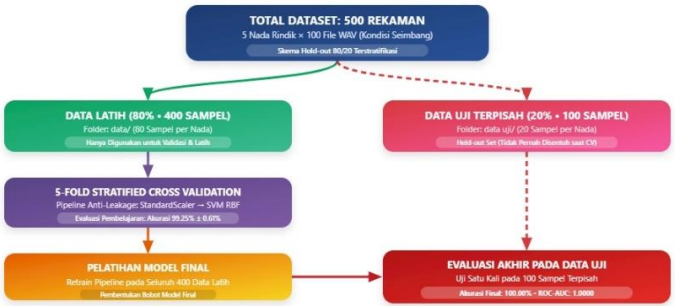


Figure 4. Evaluation protocol stratified 80/20 hold out split

Table 4. Distribution of the 500 recordings between the training and held out test subsets (stratified 80/20 hold out).

Note	*Training subset*	*Test subset*	*Total*
Dang	80	20	100
Deng	80	20	100
Ding	80	20	100
Dong	80	20	100
Dung	80	20	100
Total	**400**	**100**	**500**

Data leakage occurs when information from outside the training data influences the model, which inflates the estimated performance [22], [23]. Three safeguards were applied. First, the training and test subsets are stored in physically separate folders, and the test subset is never read during cross validation, tuning, or model selection. Second, the application computes an MD5 hash for every file and compares the hashes across the two subsets no duplicate was found among the 400 training and 100 test files, so the two subsets are disjoint at file level (Figure 5). Third, all feature scaling and classification steps are encapsulated in a scikit learn Pipeline [25]. Within each fold of the 5 fold stratified cross validation, StandardScaler.fit() is invoked only on the training partition of that fold during validation and final testing, only StandardScaler.transform() is applied, so no information from held out data contaminates the scaling statistics.



Figure 5. MD5 duplicate verification between the training folder and the test folder

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

MD5 verification detects byte identical files only. It cannot detect recordings that are acoustically similar because they were captured consecutively in the same session this residual risk is discussed in Section 3.7

3. RESULTS AND DISCUSSION

This section covers signal analysis, model training, classification performance under the two stage protocol, the ablation study, the baseline comparison, web deployment, and a discussion of limitations. Unless stated otherwise, cross validation figures refer to the 400 sample training subset and test figures refer to the 100 held out samples.

3.1. Signal Analysis and Feature Representation

Each audio sample is represented as a 16 dimensional feature vector derived from four DWT sub bands (A3, D3, D2, D1), where each sub band contributes four statistical features mean, standard deviation, energy, and entropy. This forms a 400 × 16 training matrix and a 100 × 16 test matrix (500 × 16 overall). RMS normalization is intended to stabilize signal energy across different striking intensities and to produce more consistent amplitude patterns across the sub bands. Figure 5 compares the original audio signal with the RMS normalized wavelet coefficients and illustrates how this step reduces the amplitude spread of the coefficients. Whether this visual stabilization translates into a measurable gain in classification accuracy is examined in Section 3.4. Figure 6 shows the same comparison as generated by the application for a Dang sample, and Figure 7 shows one recording of each note.

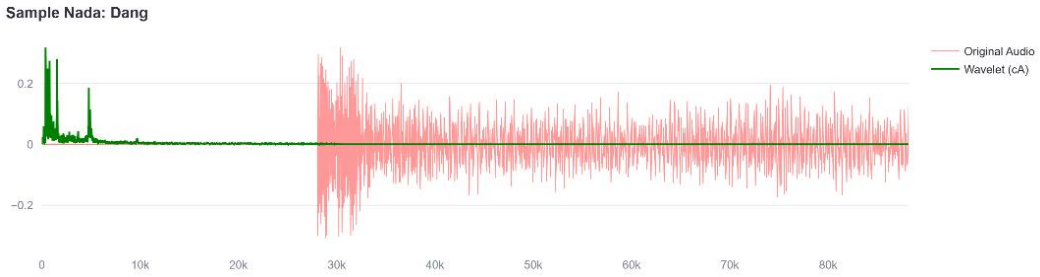


Figure 6. Visual comparison between the original audio signal and RMS normalized wavelet coefficients

Figure 6 shows the comparative visualization produced by the feature extraction tab for a Dang sample (Dang (1).wav). The upper panel plots the waveform, sampled every 100 points, with and without RMS normalization. The raw recording has a very low amplitude, whereas the RMS normalized signal spans roughly ±0.2, so the overall signal level entering the wavelet stage is brought to a common scale regardless of how hard the blade was struck. The lower panel compares the level 3 db1 approximation coefficients with RMS

Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1142

normalization the leading coefficients reach magnitudes of about 400 or more, whereas without it they remain an order of magnitude smaller. The normalization therefore changes the scale of the wavelet coefficients substantially. Because the same RMS scaling is applied to every recording, it should reduce amplitude differences between recordings. Whether this also improves classification is tested in the ablation study (Section 3.4).

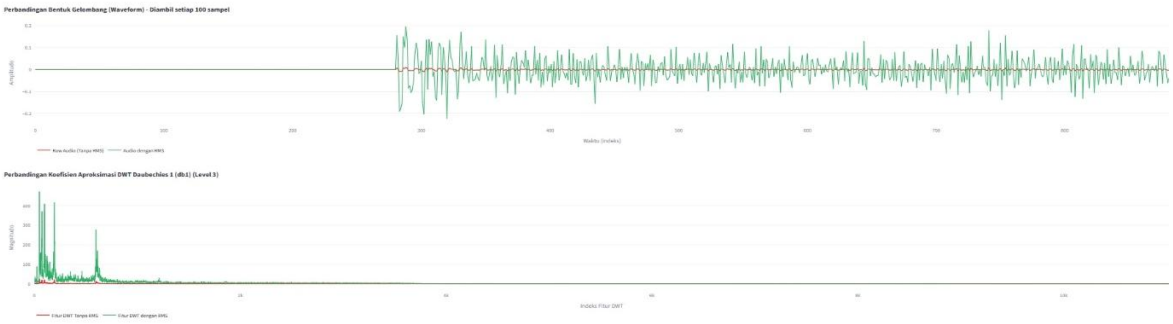


Figure 7. Effect of RMS normalization for a Dang sample

Figure 7 shows one recording of each note, with the original waveform (red) and the level 3 approximation coefficients (green). The notes differ in their temporal structure. The strike onset occurs at about 28,000 samples in the Dang, Ding, and Dung examples but at about 41,000–43,000 samples in the Deng and Dong examples. The Deng and Dong recordings show amplitude modulated envelopes that end at about 62,000 - 63,000 samples, whereas the Dang, Ding, and Dung recordings keep a low level tail until the end of the 2 second window. The leading approximation coefficients also differ in the position and number of their dominant peaks. These differences are visible before any classification and support the use of time frequency features. The variable onset time also illustrates why silence trimming is applied before decomposition (Section 2.3).

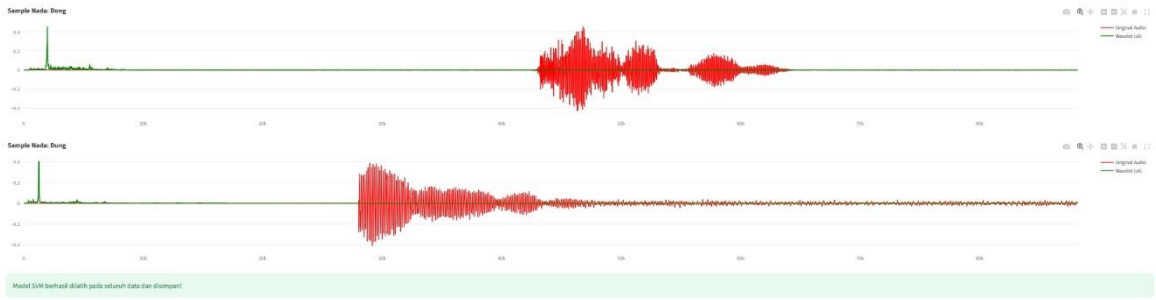


Figure 8. Original waveform and level 3 approximation coefficients (cA) of one sample

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

3.2. SVM Model Training and Optimization

During the training phase, SVM parameter optimization was conducted using GridSearchCV with 5 fold stratified cross validation on the 400 sample training subset. The search identified C = 10 and gamma = 'auto' as the optimal configuration, with a best mean cross validation accuracy of 99.25%. The application offers three training modes (Figure 8): standard training on all training data with C = 1.0 and gamma = 'scale', k fold cross validation without tuning, and GridSearchCV auto tuning. The last mode produced the reported results, and the trained model is stored as a .dill file for the identification tab. The 'auto' setting scales the kernel coefficient by the number of features, which suited the spread of the feature values in this dataset. Feature scaling and classification were executed inside the leakage controlled pipeline described in Section 2.6, so the scaler was fitted on training partitions only. Other classification approaches, such as Decision Tree and Hidden Markov Model (HMM), have been explored in previous works on audio recognition [7], [9] they were not benchmarked in this study, and therefore no claim is made about their performance relative to SVM.

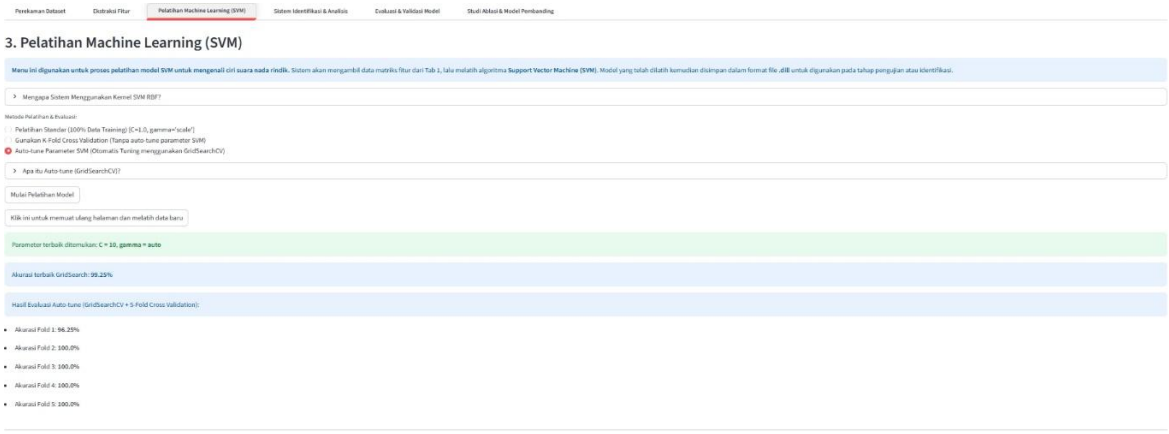


Figure 9. Training tab with GridSearchCV auto tuning: best parameters

3.3. Classification Performance and Statistical Validation

The system was evaluated under the two stage protocol described in Section 2.6. In the first stage, 5 fold stratified cross validation was applied exclusively to the 400 sample training subset. The mean validation accuracy across the five folds was 99.25%, with a standard deviation of 0.61% (population standard deviation of the five fold scores Table 5). This figure characterizes learning stability and was used for hyperparameter tuning it is not the final performance of the system on unseen data. The narrow range of the fold scores (98.75% - 100.00%) suggests that the model is not strongly overfitted to any particular part of the training data. In the second stage, the model was retrained on all 400 training samples and evaluated exactly once on the 100 held out test samples. The held out accuracy was

100.00%, with precision, recall, and F1 score of 100.00% and a ROC AUC of 1.0000 (Table 7). Tables 5 and 6 therefore document training phase validation, whereas Table 7, Figure 12, and Figure 13 constitute the definitive performance evidence.

Table 5. Per fold accuracy during 5 fold stratified cross validation on the training subset (400 samples training phase result, not the held out test result).

Validation fold	*Accuracy (%)*
Fold 1	98.75
Fold 2	100.00
Fold 3	98.75
Fold 4	98.75
Fold 5	100.00
Mean	99.25
Standard deviation	0.61

Table 5 uses the stratified fold assignment of the evaluation tab. The training tab (Figure 8) builds its own five folds, which gave per fold accuracies of 96.25%, 100.00%, 100.00%, 100.00%, and 100.00% (population standard deviation 1.50%). Both fold assignments give the same mean accuracy of 99.25%, which corresponds to three misclassified samples out of 400 (397/400), but they distribute those errors differently across the folds, so the per fold values and the standard deviation differ (0.61% in Table 5 versus 1.50% here). With so few errors, the standard deviation is sensitive to the fold assignment and should not be over interpreted. Table 6 and Figure 9 give the per note view of this training phase evaluation over the 400 training samples (accuracy 99.25%, macro precision 99.28%, recall 99.25%, and F1 score 99.25%). Three samples, one Ding and two Dung, were assigned to Dang, which lowers the precision of Dang to 0.96. The cause was not investigated. Overlapping harmonic content between these notes is a plausible but unverified explanation. No such errors occurred on the held out subset (Table 7).

Table 6. Training phase classification report on the 400 training samples

Note	*Precision*	*Recall*	*F1 score*	*Support*
Dang	0.96	1.00	0.98	80
Deng	1.00	1.00	1.00	80
Ding	1.00	0.99	0.99	80
Dong	1.00	1.00	1.00	80
Dung	1.00	0.97	0.99	80
Accuracy	0.99	0.99	0.99	400
Macro average	0.99	0.99	0.99	400
Weighted average	0.99	0.99	0.99	400

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id



Figure 10. Confusion matrix of the training phase evaluation on the 400 training samples.

Figure 10 shows how the evaluation tab presents the two stages separately. GridSearchCV selected C = 10 and gamma = 'auto' with a best cross validation score of 99.25%. The five fold scores and their mean (99.25% ± 0.61%) are displayed in a box that is explicitly labeled as an indicator for training and tuning, not as the final test result, whereas the held out result (100.00% accuracy, precision, recall, and F1 score. ROC AUC 1.0000) is displayed and labeled as the final figure of the study. The classification report, confusion matrix, and ROC curves that follow (Table 7, Figure 12, and Figure 13) are computed from the same 100 held out samples

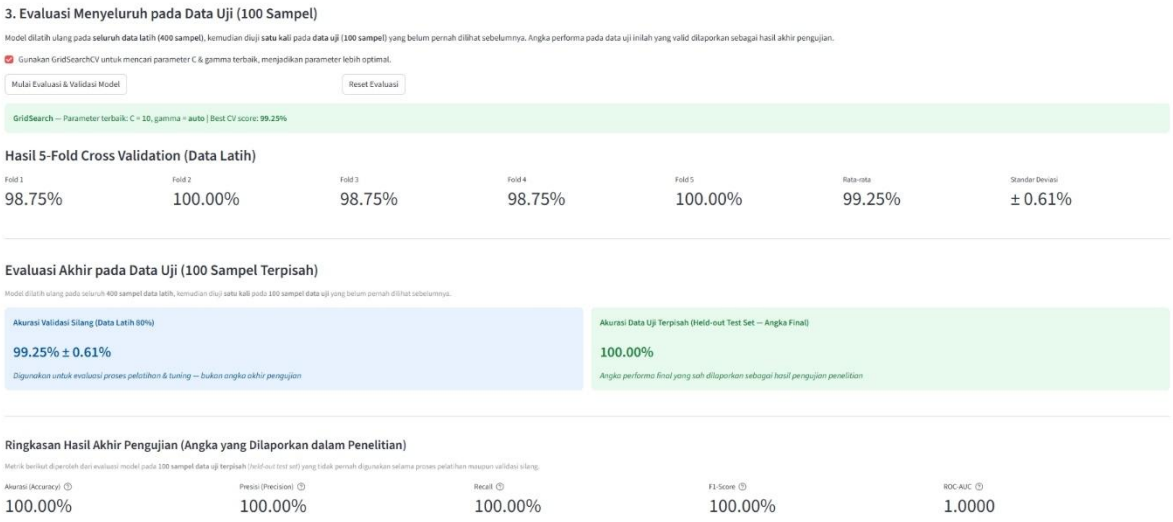


Figure 11. Evaluation tab GridSearchCV result, 5 fold cross validation on the training subset, and the final evaluation on the held out test subset.

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1142

Table 7. Classification report on the held out test subset (100 samples, 20 per note).

Note	*Precision*	*Recall*	*F1 score*	*Support*
Dang	1.00	1.00	1.00	20
Deng	1.00	1.00	1.00	20
Ding	1.00	1.00	1.00	20
Dong	1.00	1.00	1.00	20
Dung	1.00	1.00	1.00	20
Accuracy	1.00	1.00	1.00	100
Macro average	1.00	1.00	1.00	100
Weighted average	1.00	1.00	1.00	100



Figure 12. Confusion matrix of the SVM classifier on the held out test subset (100 samples).

Figure 12 displays the confusion matrix on the held out test subset. All 20 samples of each note lie on the main diagonal, and the misclassification analysis of the application reported no wrongly predicted sample. The one vs rest ROC curves in Figure 13 show an AUC of 1.0000 for all five notes, meaning that the classifier scores always ranked the correct note above the alternatives on this test subset. A perfect result must nevertheless be interpreted with care. With only 20 test samples per note, one performer, one instrument, and a quiet recording environment, it indicates that the five notes are well separated in the 16 dimensional feature space under controlled conditions it does not show that the model would generalize to other performers or noisy environments. The fold scores below 100% during cross validation (Table 5) and the training phase confusion matrix (Figure 10) also show that occasional single sample errors did occur during training phase validation.

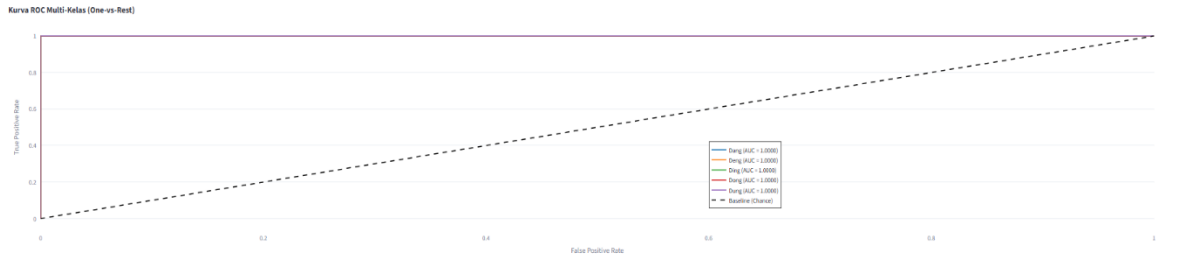


Figure 13. One vs rest ROC curves of the five notes on the held out test subset (AUC = 1.0000 for every note).

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Compared with earlier Rindik work that relied on the Fast Fourier Transform (FFT) [1], which discards temporal signal changes and struggles with the transient nature of bamboo strikes, the DWT based representation retains time frequency information [2], [11], [14]. In our own comparison (Section 3.5), the FFT based baseline reached only 85.00% test accuracy, which is consistent with this limitation. Applying DWT without energy normalization may leave the feature vectors sensitive to amplitude variability caused by inconsistent striking forces the sub band RMS normalization was designed to address this issue, although its measurable benefit on the present dataset is not statistically significant (Section 3.4).

3.4. Ablation Study on RMS Normalization

To isolate the contribution of each pipeline component, three scenarios were evaluated with the same SVM RBF pipeline, the same 5 fold stratified cross validation on the training subset, and the same held out test subset. Scenario A (no DWT) uses only time domain statistical features (mean, standard deviation, maximum, minimum, RMS, zero crossing rate, and energy) extracted directly from the audio window. Scenario B (DWT without normalization) applies the level 3 db1 decomposition but extracts the statistics from the raw, non normalized sub band coefficients. Scenario C (DWT with RMS normalization) is the full proposed pipeline. Table 8 reports the results, and Table 9 reports a paired t test ($\alpha = 0.05$) on the per fold cross validation scores, which are listed in Table 10.

Table 8. Ablation study results (5 fold cross validation on the training subset and a single evaluation on the held out test subset).

Method	*CV accuracy (%)*	*Mean ± SD*	*Held out test accuracy (%)*
No DWT (time domain statistical features)	83.75	± 4.74	80.00
DWT without normalization	99.50	± 0.61	100.00
DWT + RMS normalization (proposed)	99.25	± 0.61	100.00

Table 9. Paired t test on per fold cross validation accuracy ($\alpha = 0.05$).

Comparison	*t statistic*	*p value*	*Conclusion*
DWT + RMS vs. no DWT	+7.207	0.002	Significant
DWT + RMS vs. DWT without normalization	−1.000	0.374	Not significant

Table 10. Per fold cross validation accuracy (%) of all feature configurations on the training subset.

Fold	*No DWT*	*DWT without normalization*	*DWT + RMS (proposed)*	*FFT + SVM*	*MFCC + SVM*
Fold 1	85.00	98.75	98.75	93.75	100.00
Fold 2	85.00	100.00	100.00	95.00	98.75
Fold 3	77.50	98.75	98.75	97.50	98.75
Fold 4	80.00	100.00	98.75	96.25	98.75
Fold 5	91.25	100.00	100.00	96.25	100.00
Mean	83.75	99.50	99.25	95.75	99.25

DWT decomposition clearly matters. Removing it lowers the mean cross validation accuracy from 99.25% to 83.75% and the held out accuracy from 100.00% to 80.00%, and the paired difference is significant (t = +7.207, p = 0.002) the per fold scores of Scenario A were 85.00%, 85.00%, 77.50%, 80.00%, and 91.25%. In contrast, adding sub band RMS normalization to the DWT features did not produce a measurable gain on this dataset. Scenario B scored slightly higher in cross validation (99.50% vs. 99.25%, a difference of a single validation sample in fold 4) and identically on the held out subset, and the paired difference is not significant (t = −1.000, p = 0.374). The present data therefore do not allow us to conclude that RMS normalization improves accuracy. Both DWT variants operate close to the ceiling, which leaves little room for any gain to appear, and evidence for the benefit of the normalization would require data in which the striking force and the recording conditions are deliberately varied. Finally, because the five fold scores are derived from overlapping training sets and are not independent, the p values should be read as indicative rather than definitive [27], [28]. Figure 14 plots the cross validation and held out accuracy of the three scenarios.

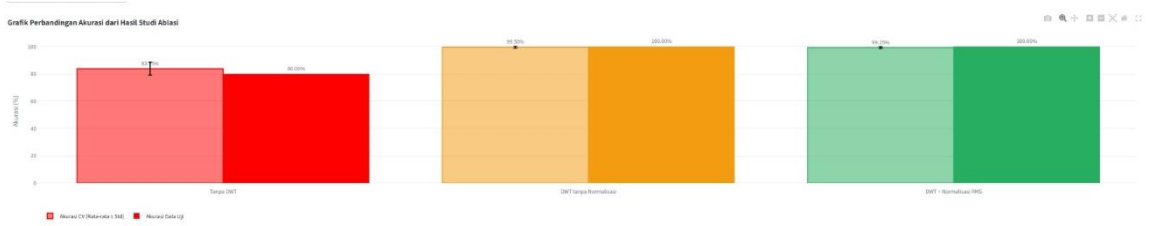


Figure 14. Ablation study cross validation accuracy and held out test accuracy for the three scenarios

3.5. Comparative Analysis with Existing Methods

Table 11 benchmarks the proposed method against FFT based and MFCC based feature extraction. All four methods use the same SVM RBF classifier inside a scikit learn Pipeline with StandardScaler, the same 5 fold stratified cross validation on the 400 training samples, and a single evaluation on the 100 held out test samples. Precision, recall, and F1 score are weighted averages. FFT + SVM achieved the lowest test accuracy (85.00%, F1 score 85.06%), consistent with the inability of a purely frequency domain representation to capture transient, non stationary information. MFCC + SVM, DWT + SVM (without RMS), and the proposed DWT + RMS + SVM pipeline all reached 100.00% on the held out test subset. On this controlled dataset, the proposed method is therefore clearly better than the FFT baseline but cannot be distinguished in accuracy from the MFCC baseline or from DWT without normalization. MFCC was originally designed for speech [16], and its spectral smoothing may be less suitable for impulsive percussive signals, but no accuracy disadvantage was observed here. The potential advantage of the proposed pipeline lies in its designed amplitude invariance and compact 16 dimensional feature vector, which

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

remains a hypothesis to be tested on more challenging data. Figure 15 plots the accuracies of the four methods, and their per fold scores are listed in Table 10

Table 11. Accuracy Comparison of Feature Extraction Methods

Method	*CV accuracy (%)*	*Mean ± SD*	*Test accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1 score (%)*
FFT + SVM	95.75	± 1.27	85.00	86.86	85.00	85.06
MFCC + SVM	99.25	± 0.61	100.00	100.00	100.00	100.00
DWT + SVM (without RMS)	99.50	± 0.61	100.00	100.00	100.00	100.00
DWT + RMS + SVM (proposed)	99.25	± 0.61	100.00	100.00	100.00	100.00

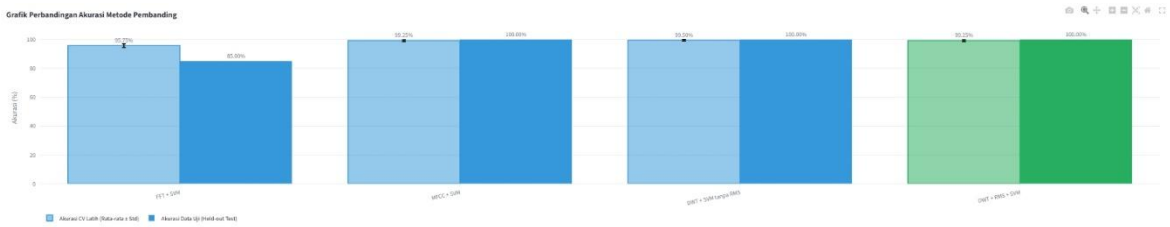


Figure 15. Cross validation accuracy and held out test accuracy of the four feature extraction methods

3.6. Web Implementation and Real time Testing

The trained model was integrated into a web interface using Streamlit (Python). Implementing the system as a web application demonstrates the practical use of machine learning in real world cultural preservation scenarios, supported by modern computing and Python data analysis libraries [17], [18]. The application is organized into six functional tabs. (Tab 1) Dataset Recording captures monophonic audio samples from a condenser microphone and organizes them by note class (Figure 3). (Tab 2) Feature Extraction visualizes the effect of RMS normalization on a selected sample (Figure 7) and batch processes the training recordings with the level 3 db1 DWT and an optional RMS normalization switch, saving a 400 × 16 feature matrix (Figure 16). (Tab 3) SVM Model Training trains the SVM RBF classifier in one of three modes, including GridSearchCV auto tuning, and stores the model (Figure 9). (Tab 4) Identification and Analysis accepts uploaded .wav files or live microphone recordings and returns the predicted note with a confidence score and a wavelet coefficient visualization (Figure 17). (Tab 5) Model Evaluation and Validation implements the two stage protocol of Section 2.6, including the MD5 duplicate check, 5 fold cross validation, the held out evaluation, the confusion matrix, the ROC curves, and downloadable CSV reports and (Tab 6) Ablation Study and Baseline Comparison runs the ablation scenarios, the paired t-test, and the FFT and MFCC baselines on the actual dataset. Figure 17 shows the identification tab, where users can upload an audio file or record live, optionally apply RMS normalization, and instantly see the predicted note together with the confidence of every class.

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1142

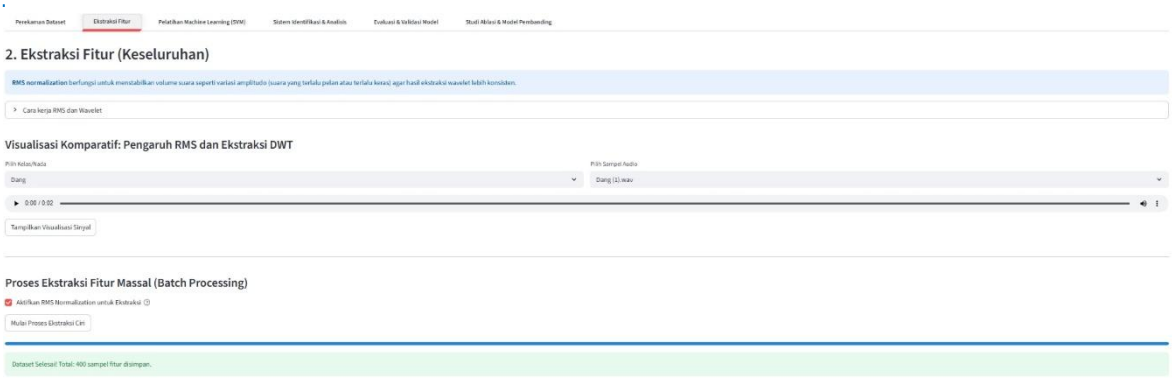


Figure 16. Feature extraction tab comparative visualization, batch extraction with the RMS normalization.

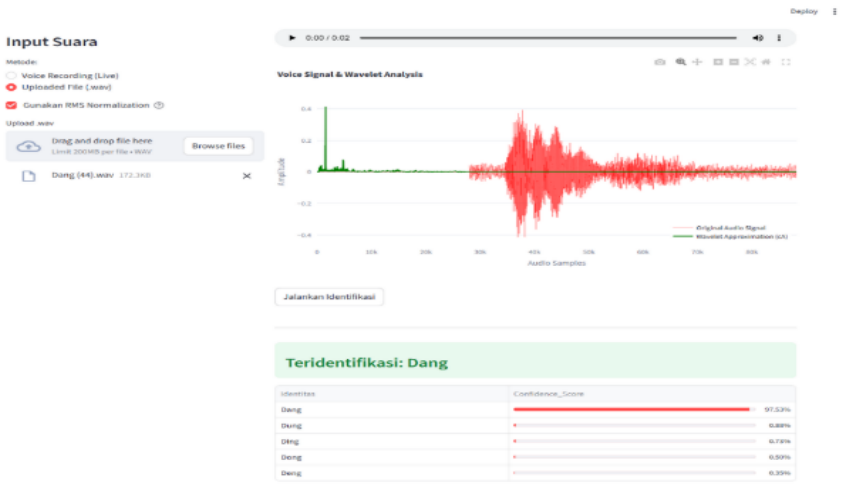


Figure 17. Web application interface for audio testing and prediction

During testing, the model returned a clear top prediction for representative samples of all five note classes, with highest confidence scores between 89.36% and 97.83% (Table 12 and Figure 17). In the example of Figure 17, Dang (99).wav was identified as Dang with 89.36% confidence, and the remaining probability was spread over Dung (5.23%), Ding (3.86%), Deng (0.91%), and Dong (0.64%). Averaged over the four non target classes, the probability stayed below 3% in every case in Table 12, although a single non target class can occasionally reach about 5%, as Dung does in this example. These confidence values come from representative recordings and are not a substitute for the full held out evaluation in Section 3.3. They also show that some individual recordings are classified with a smaller margin than the accuracy figures alone suggest.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Table 12. Model confidence score per note class

Test Data	Identified Note	*Highest Confidence Score*	Average Probability of Other Classes
Dang Note	Dang	97.47%	< 1.00%
Deng Note	Deng	97.33%	< 1.00%
Ding Note	Ding	92.95%	< 3.00%
Dong Note	Dong	96.93%	< 2.00%
Dung Note	Dung	97.83%	< 1.00%
Dang Note (Figure 13)	Dang	89.36%	2.66%

The balanced held out precision, recall, and F1 score (all 100.00%) indicate that the classifier treats the five notes consistently. Since no other classification algorithms were benchmarked in this study, no claim is made that SVM generalizes better than simpler algorithms.

These results, while encouraging, come with an important caveat all recordings were made under controlled conditions at a fixed distance of ±30 cm with minimal background noise. Real performance environments introduce variables such as crowd noise, polyphonic signals, overlapping sounds, and inconsistent playing styles that this setup never encountered. With only 500 samples from a single performer, the model has not been stress tested against the full range of real world variability, and its accuracy may not hold at the same level outside the laboratory. Further evaluation on a larger and more varied dataset, including live performances, is therefore recommended.

The real time implementation results demonstrate that the proposed system performs reliably under controlled conditions. However, beyond implementation performance, a deeper analysis is required to understand the factors contributing to the model's effectiveness and its limitations. Therefore, the following subsection presents a discussion of the experimental findings.

3.7. Discussion

The DWT + RMS + SVM pipeline reached a mean cross validation accuracy of 99.25% ± 0.61% on the training subset and 100.00% accuracy on the held out test subset. Two factors plausibly explain these results. First, DWT captures both temporal and spectral characteristics of non stationary bamboo signals, which FFT cannot do because it discards temporal information the ablation study confirms the importance of DWT, as removing it reduced the held out accuracy from 100.00% to 80.00% ($p = 0.002$ for the cross validation scores). Second, RMS normalization is designed to correct energy inconsistencies across sub bands caused by variable striking forces. On this dataset, however, the statistical evidence for an additional accuracy gain is lacking ($p = 0.374$), so this component should currently be regarded as a design hypothesis whose benefit still needs to be confirmed under more varied conditions.

Compared with MFCC, which was developed for speech like signals with smooth spectral envelopes [16], the DWT based approach offers time frequency localization that suits the percussive amplitude dynamics of bamboo instruments. In this study, however, MFCC + SVM reached the same held out accuracy (100.00%), so no accuracy advantage over MFCC can be claimed on this dataset. The advantage over the FFT baseline (100.00% vs. 85.00% test accuracy) is clear.

Despite these strong results, the study has clear limitations. First, the dataset consists of 500 samples recorded under controlled, low noise conditions by a single performer on a single instrument, which likely contributes to the near ceiling accuracy. Second, with 20 test samples per note, a perfect score still leaves statistical uncertainty the exact 95% (Clopper–Pearson) confidence interval for 100 correct predictions out of 100 has a lower bound of approximately 96.4%. Third, MD5 verification only detects byte identical files it cannot rule out acoustic similarity between recordings captured consecutively in the same session, which may make a random hold out split more favorable than a split by session or by performer, a known risk in the evaluation of music information retrieval systems [26]. Fourth, the paired t test rests on only five non independent fold scores. Finally, the system handles only monophonic notes and has not been tested in polyphonic or noisy environments. Future work should address these gaps by expanding dataset diversity (more performers, instruments, and recording sessions), using session level or performer level splits, deliberately varying the striking force to test the RMS normalization hypothesis, integrating noise preprocessing, and extending the model to polyphonic recognition and mobile deployment.

4. CONCLUSION

This study presented a machine learning based approach for automatic Balinese Rindik note recognition that integrates the Discrete Wavelet Transform (DWT), sub band RMS normalization, and a Support Vector Machine (SVM). Using a 16 dimensional feature vector and the parameters C = 10 and gamma = 'auto' identified via GridSearchCV, the system achieved a mean 5 fold cross validation accuracy of 99.25% ± 0.61% on the 400 sample training subset, which is a training phase result, and 100.00% accuracy, precision, recall, and F1 score (ROC AUC = 1.0000) in a single evaluation on 100 held out test samples. File level MD5 verification confirmed that the two subsets share no recordings.

The ablation study showed that DWT decomposition is the decisive component (cross validation accuracy of 83.75% without DWT vs. 99.25% with DWT and RMS normalization $p = 0.002$), whereas sub band RMS normalization did not yield a statistically significant improvement over DWT alone on this controlled dataset ($p = 0.374$). The FFT based baseline reached 85.00% test accuracy, while the MFCC based baseline matched the 100.00% of the proposed method. The contribution of this work is therefore a leakage aware, reproducible evaluation of sub band RMS normalization for Rindik notes, together with a

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

working Streamlit application whose identification confidence ranged from 89.36% to 97.83% on representative samples, rather than a demonstrated accuracy advantage over existing features.

However, this study is limited to controlled recording conditions with a relatively small dataset of 500 samples from a single performer and low ambient noise. The model's performance under real world conditions, including noisy environments, polyphonic signals, and varied playing techniques, has not yet been evaluated, and further testing in such settings is required. For future work, researchers should expand the dataset to include more diverse samples from different performers and instruments, use session level or performer level data splits, deliberately vary the striking force to test whether RMS normalization provides a measurable benefit, test the system in uncontrolled acoustic environments, and add polyphonic note recognition capabilities (multiple blades struck simultaneously). Incorporating noise cancellation techniques would improve robustness. Deploying the trained model in a mobile application would also make it easier for artists and the public to use it for cultural preservation.

REFERENCES

[1] I. K. K. Saduadnyana and I. K. G. Suhartana, "Implementasi double frequency modulation dan FFT dalam sintesis suara Rindik," JELIKU (Jurnal Elektronik Ilmu Komputer Udayana), vol. 11, no. 1, pp. 93–100, 2022, doi: 10.24843/JLK.2022.v11.i01.p10.

[2] Y. Christian and I. D. M. B. A. Darmawan, "Wavelet transformation and spectral subtraction method in performing automated Rindik song transcription," Jurnal Ilmu Komputer dan Informasi, vol. 15, no. 1, pp. 9–19, 2022, doi: 10.21609/jiki.v15i1.1009.

[3] I. G. Ismanmadi, "Perbandingan kinerja algoritma K NN dan SVM dalam sistem klasifikasi genre musik gamelan Bali," Jurnal Ilmiah Teknologi dan Bisnis (ITBI), vol. 8, no. 1, pp. 1–9, 2023, doi: 10.51211/itbi.v8i1.2417.

[4] T. B. Putri, S. Saidah, B. Hidayat, F. Qothrunnada, and D. Darwindra, "Deteksi emosi berdasarkan sinyal suara manusia menggunakan Discrete Wavelet Transform (DWT) dengan klasifikasi Support Vector Machine (SVM)," Jurnal Ilmu Komputer dan Informatika, vol. 3, no. 1, pp. 1–10, 2023, doi: 10.54082/jiki.45.

[5] Y. Y. Septian and Y. Suyanto, "Identifikasi genre musik menggunakan SVM dan ekstraksi ciri MFCC," Indonesian Journal of Electronics and Instrumentation Systems (IJEIS), vol. 12, no. 1, pp. 1–10, 2022, doi: 10.22146/ijeis.70898.

[6] S. K. S. Ramadhani, G. Budiman, and B. Hidayat, "Deteksi nada dasar alat musik Panting menggunakan Compressive Sensing dengan MFCC dan SVM," Journal of Electrical System and Control Engineering (JESCE), vol. 6, no. 2, pp. 1–10, 2023, doi: 10.31289/jesce.v6i2.8338.

[7] F. Qothrunnada, S. Saidah, B. Hidayat, T. B. Putri, and D. Darwindra, "Deteksi nada pada alat musik teranika menggunakan Discrete Wavelet Transform dengan klasifikasi Decision Tree," Jurnal Teknologi Informasi dan Komputer, vol. 4, no. 2, pp. 570–578, 2023, doi: 10.52436/1.jutif.2023.4.2.570.

[8] I. G. A. N. Diputra Wiraguna and L. A. A. Rahning Putri, "Klasifikasi genre musik menggunakan Support Vector Machine berdasarkan spectral features," Jurnal Nasional Teknologi Informasi dan Aplikasinya, vol. 1, no. 3, pp. 20–28, 2023, doi: 10.24843/JNATIA.2023.v01.i03.p20.
[9] S. Hidayat, A. S. Anas, S. A. A. Yusuf, and M. Tajuddin, "Sistem pengenalan pembicara dengan metode wavelet-MFCC dan pengklasifikasi Hidden Markov Models (HMM)," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 1, pp. 1–8, 2021, doi: 10.25126/jtiik.0813284.
[10] I. K. R. A. Putra, L. A. A. Rahning Putri, and I. M. Widiartha, "Klasifikasi gamelan Bali menggunakan Support Vector Machine (SVM)," JELIKU (Jurnal Elektronik Ilmu Komputer Udayana), vol. 13, no. 2, pp. 1–10, 2024, doi: 10.24843/JLK.2024.v13.i02.p22.
[11] J. Prasetya, A. S. Syahputri, and M. D. A. Bahtiar, "Pengolahan sinyal suara menggunakan transformasi wavelet untuk monitoring kondisi bearing motor induksi," Jurnal Alinier, vol. 5, no. 2, pp. 1–10, 2024, doi: 10.36040/alinier.v5i2.10003.
[12] M. I. Fauzi, R. Magdalena, and B. Hidayat, "Deteksi akor dan melodi pada file WAV gitar fingerstyle menggunakan metode DWPT dan K-NN," Journal of Electrical, Electronic, System, and Control Engineering (JESCE), vol. 3, no. 2, pp. 1–8, 2020, doi: 10.31289/jesce.v3i2.3324.
[13] A. Eviyanti, A. S. Fitrani, and U. K. Nisak, "Ekstraksi fitur rekaman sinyal EEG menggunakan metode Discrete Wavelet Transform (DWT) untuk klasifikasi iktal epilepsi," Prosiding Seminar Nasional Psikologi dan Sosial Humaniora, vol. 3, pp. 231–238, 2022, doi: 10.21070/pssh.v3i.231.
[14] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674–693, 1989, doi: 10.1109/34.192463.
[15] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.
[16] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357–366, 1980, doi: 10.1109/TASSP.1980.1163420.
[17] M. Suryati, T. A. D. Saputra, and I. H. Ikasari, "Penggunaan Python dalam analisis data dengan machine learning," JRIIN: Jurnal Riset Informatika dan Inovasi, vol. 2, no. 10, pp. 1959–1968, 2025.
[18] A. Tholib, Implementasi Algoritma Machine learning Berbasis Web dengan Framework Streamlit. Pustaka Nurja, LP3M Universitas Nurul Jadid, 2023.
[19] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in Proc. 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, 2003, pp. 282–289, doi: 10.1145/860435.860487.
[20] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, 2002, doi: 10.1109/TSA.2002.800560.
[21] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," IEEE Signal Processing Magazine, vol. 36, no. 1, pp. 20–30, 2019, doi: 10.1109/MSP.2018.2869928.
[22] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 4, Art. 15, 2012, doi: 10.1145/2382577.2382579.
[23] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, Art. 100804, 2023, doi: 10.1016/j.patter.2023.100804.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

[24] G. C. Cawley and N. L. C. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," Journal of Machine Learning Research, vol. 11, pp. 2079–2107, 2010.
[25] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[26] B. L. Sturm, "A simple method to determine if a music information retrieval system is a "horse"," IEEE Transactions on Multimedia, vol. 16, no. 6, pp. 1636–1644, 2014, doi: 10.1109/TMM.2014.2330697.
[27] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," Neural Computation, vol. 10, no. 7, pp. 1895–1923, 1998, doi: 10.1162/089976698300017197.
[28] C. Nadeau and Y. Bengio, "Inference for the generalization error," Machine Learning, vol. 52, no. 3, pp. 239–281, 2003, doi: 10.1023/A:1024068626366.