Comparative Analysis of Random Forest and Support Vector Machine for Sundanese Dialect Classification Using Speech Recognition Features

Abdull Halim Anshor[1]*, Tri Ngudi Wiyatno[2]
Department of Informatics Engineering[1], Department of Industrial Engineering[2]
Universitas Pelita Bangsa
Bekasi, Indonesia
abdulhalimanshor@pelitabangsa.ac.id[1], tringudi@pelitabangsa.ac.id[2]

Abstract— This study investigates the classification of West and South Sundanese dialects using Random Forest (RF) and Support Vector Machine (SVM). Using a dataset of 100 recordings with features extracted via Mel Frequency Cepstral Coefficient (MFCC), models were evaluated by accuracy, precision, recall, and F1-score. Results show RF achieved an accuracy of 93.33%, outperforming SVM's 73.33%. The analysis demonstrates that RF is more reliable in distinguishing dialectal features. This research contributes to regional speech recognition, supporting language preservation and improved dialectal analysis.

Keywords—Classification of Sundanese Dialects, Machine Learning, Random Forest, Support Vector Machine, Mel-Frequency Cepstral Coefficient

I. INTRODUCTION

Machine learning (ML) and speech recognition technologies have advanced significantly in recent years, impacting numerous sectors such as linguistics, healthcare, and information systems [1], [2], [3]. The ability to recognize dialects accurately plays an important role in preserving regional languages and improving personalized speech applications [4], [5]. However, the focus of most speech recognition systems remains on major languages, often neglecting the acoustic complexity of regional dialects like West and South Sundanese [6], [7].

The limitation of existing models to differentiate subtle variations among dialects leads to decreased inclusivity and accuracy in voice-driven systems [8], [9]. In previous studies, Random Forest (RF) and Support Vector Machine (SVM) algorithms have shown effectiveness in various classification tasks. RF has been successfully used for sentiment analysis [10], land use classification [7], speech emotion recognition , and English dialect identification [1]. RF models demonstrate advantages in handling high-dimensional and nonlinear datasets [11], [12], which are common in speech and audio analysis [13], [14].

Support Vector Machine (SVM), on the other hand, has been applied for tasks such as bird species audio classification , gender recognition from voice [14], and emotion expression detection in speech [15]. Studies have indicated that SVM can achieve high accuracy in well-structured datasets, although it may struggle with highly complex patterns compared to ensemble methods like RF [8], [16].

Hybrid and optimized models have also emerged to enhance classification accuracy. For example, CNN-Attention-Optimized RF models have been developed for detecting abusive tweets [5], and multimodal machine learning has been applied to detect markers of mental health through speech [9]. Feature extraction methods, particularly Mel-Frequency Cepstral Coefficients (MFCC), are commonly employed in speech analysis for their ability to capture critical acoustic features [17], [18], [19].

The availability of Sundanese speech datasets [13], opens opportunities for studying regional dialects. However, existing research has focused more on disease detection [11], [16], urban land use mapping [17], action recognition [15], and sentiment analysis [20], leaving a gap in the classification of Sundanese dialects using RF and SVM approaches.

Furthermore, studies have highlighted the importance of adapting classification systems to specific acoustic patterns in regional languages [21], [22], [23]. The combination of feature selection and ensemble models has shown potential for improving classification performance in complex datasets, including speech impairments and voice disorders [24], [22].

Although comparative studies between RF and SVM have been conducted in other domains such as medical diagnostics [24], tree species classification [25], risk analysis in peer-to-peer lending [26], and judicial decision predictions [27], research directly comparing the two algorithms for Sundanese dialect classification remains scarce.

This research contributes to filling this gap by directly comparing RF and SVM in Sundanese dialect classification, which has not been extensively explored in past studies. The novelty lies in the application of MFCC features on local dialect audio combined with classical ML classifiers. Therefore, this study aims to fill this research gap by conducting a comparative analysis of the Random Forest and Support Vector Machine algorithms, using MFCC feature extraction, for the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

classification of West and South Sundanese dialects.

II. RESEARCH METHODOLOGY

This study implements the Random Forest (RF) and Support Vector Machine (SVM) algorithms for the classification of limited sources of discussion of West and South Sundanese dialects. For this reason, this study is divided into a number of stages consisting of data collection, sound feature actracy, data sharing, model training, and model evaluation using relevant metrics [1] [2]. The research stages are shown in Figure 1.

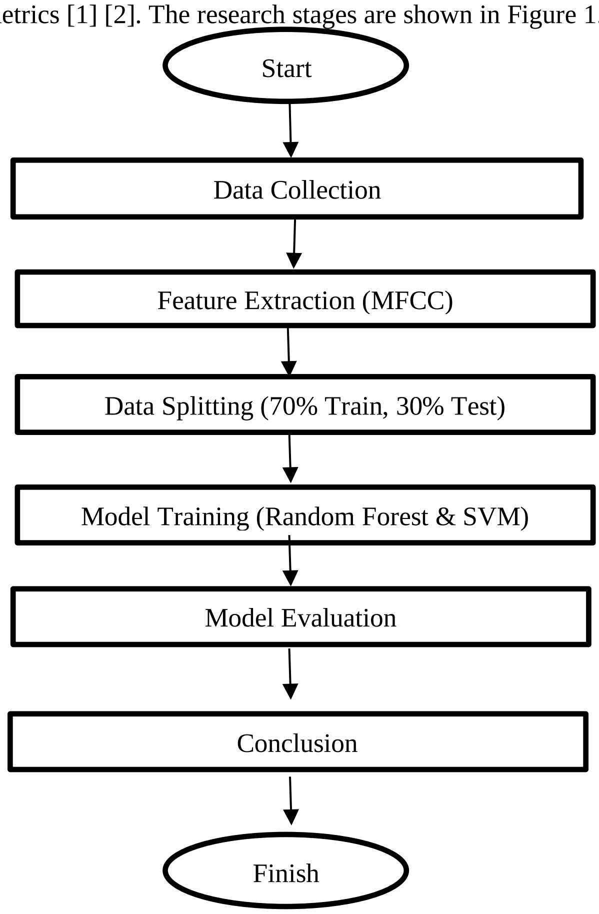

Fig. 1. The research stages

The studied literature has been categorized into three subject categories for clarity: (1) RF/SVM in audio classification, (2) MFCC in dialect detection, and (3) speech recognition in low-resource languages.

A. RF and SVM in Audio Classification

Random Forest (RF) and Support Vector Machine (SVM) are extensively employed in audio classification endeavors. Random Forest (RF) has demonstrated considerable accuracy and resilience in tasks such as dialect identification and emotion detection, attributed to its ensemble characteristics and capacity to handle nonlinear data [1], [18]. SVM demonstrates efficacy with smaller, well-organized datasets and has been employed in applications such as avian species identification, and gender categorization from audio . Both models function as robust baselines for dialect classification [2], [14].

B. MFCC in Dialect Identification

The Mel-Frequency Cepstral Coefficients (MFCC) methodology is a preeminent feature extraction method in speech analysis, esteemed for its capacity to emulate the human auditory system. It catches acoustic subtleties, rendering it especially proficient in dialect and accent categorization [11], [17]. MFCC has been extensively utilized in applications like emotion recognition, speaker verification, and dialect profiling.

C. Speech Recognition in Resource-Scarce Languages

Sundanese is classified as a low-resource language, and current studies suggest that conventional machine learning models such as Random Forest (RF) and Support Vector Machines (SVM) may surpass deep learning models in scenarios with minimal data [9], [21]. Previous research underscores the necessity of employing efficient algorithms and strong feature extraction methods in environments characterized by limited annotated data and significant dialectal complexity.

D. Data Collection

The Sundanese ASR dataset, available at Kaggle, contains a collection of Sundanese-language audio recordings accompanied by their corresponding text transcriptions. This dataset serves as a reliable resource for speech recognition tasks, particularly in dialect classification.

In this study, the data was divided into 70% for training and 30% for testing. This split was chosen to ensure balanced model evaluation, helping to maintain validity while minimizing the risk of overfitting [6].

For the training process, Random Forest (RF) and Support Vector Machine (SVM) algorithms were implemented using features extracted through Mel-Frequency Cepstral Coefficients (MFCC) [7]. The performance of both models was assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score. Additionally, ROC curves and confusion matrices were employed to provide deeper insights into classification effectiveness and error distribution [1].

E. Feature Extraction with Mel-Frequency Cepstral Coefficients (MFCC)

MFCC is a crucial technique in speech signal processing used to extract the key features of spoken audio. The process begins with pre-emphasis, which amplifies high frequency components, followed by segmenting the signal into frames and applying a Hamming window to ensure smooth transitions between segments. Next, the Fast Fourier Transform (FFT) converts the signal from the time domain into the frequency domain. The Mel scale is then applied to adjust the frequency representation in a way that aligns with human auditory perception. Finally, the Discrete Cosine Transform (DCT) reduces the spectral information into a set of essential coefficients that are efficient for the classification process. [2], [14].

.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

F. Model Training

a. Random Forest (RF)

Random Forest (RF) is an ensemble-based machine learning algorithm that builds multiple decision trees to improve classification accuracy. Each tree is trained using a random subset of the data and features, thus minimizing the risk of overfitting. Two important parameters in RF are the number of trees (n_trees) and the maximum depth of the tree (max_depth), which are usually adjusted through an optimization process [1], [14].

The RF model combines the results of a number of trees to produce a final prediction. The final prediction is calculated by the decision of combining each tree with an average or majority vote, which can be expressed by equation 1.

$$\hat{y} = \frac{1}{n\_trees} \sum_{i+1}^{n\_trees} Ti(x) \quad (1)$$

where Ti(x) is the prediction from the i-th tree to the input x. The main parameters of RF are the number of trees (n_trees) and the maximum depth (max_depth), which are set to achieve the best performance [1]. RF has advantages in handling complex data and often shows better performance compared to other algorithms on non-linear data, as proven in a number of previous studies [9].

b. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a classification algorithm that works by determining the best hyperplane to separate two groups of data in a high-dimensional space. This algorithm is known to be reliable in handling linear and non-linear data, thanks to its ability to utilize kernel functions. In this context, linear kernels and Radial Basis Function (RBF) are used to test the performance of SVM on voice data in various dialects. [21]. The decision function in the SVM method can be explained through Equation 2

$$f(x) = sign\left(\sum_{i-1}^{N} \alpha iyiK(xi,x) + b\right) \quad (2)$$

Where:
α i is the coefficient obtained during the training process and is only evaluated as zero For the support vector,
yi is the class label of the training data,
K(xi,x) is the kernel function that measures the similarity between the training data xi and the input bias x,
B For the Radial Basis Function (RBF) kernel, The kernel function K(xi,x) is defined by equation 3.

$$K(xi,x) = \exp(-\gamma \parallel xi - x \parallel 2) \quad (3)$$

In the SVM algorithm, there are two main parameters that need to be set, namely the C value and gamma (γ). The C parameter plays a role in controlling the balance between the class separation margin and tolerance for misclassification, while γ regulates the extent to which a data point influences the shape of the decision boundary. In this study, both parameters were optimized with the aim of obtaining the best classification accuracy in recognizing Sundanese dialects [6]

G. Model Evaluation

In the classification process, evaluating model performance is not just a formality, but an important part so that we know how well the model reads patterns from complex data, especially if the data distribution is unbalanced. In this study, several indicators were used, ranging from accuracy to ROC-AUC. Indeed, accuracy is often the initial benchmark, but to be honest, this metric can be misleading if one class is much larger in number. Therefore, precision and recall are more reliable, especially to see how often the model actually recognizes important classes. F1-score is also a complement, because it balances the two metrics in one number. Finally, we also need a confusion matrix and ROC curve so that we can see the error map and how well the model distinguishes between West and South Sundanese dialects, especially when comparing the Random Forest and SVM algorithms.

III. Result and Discussion

This section explains the results of the voice data processing process from the West and South Sundanese dialects. The voice data was first converted into numbers using the MFCC method, so that it could be read by a computer. Now, after becoming numbers, the data was processed using two models Random Forest and SVM to help recognize the characteristics of each dialect. The goal is for the system to be able to know the difference between the voices from the two regions more accurately.

A. Feature Extraction Results Using MFCC

Using the MFCC feature extraction process, the Sundanese dialect voice signal is converted into numeric data form through several stages, namely pre-emphasis, framing, windowing, fast Fourier transform (FFT), Mel filter, and discrete cosine transform (DCT). From a total of 100 voice recordings analyzed, each produced 13 MFCC coefficient values, which were then used as input for the Random Forest and Support Vector Machine algorithms. Both models are used to distinguish the characteristics between the West and South Sundanese dialects. Details of the results are presented in Table 1.

TABLE I. Transformation Results

Reckoning Voice	MFCC1	MFCC2	MFCC3	MFCC4	MFCC5	…….	MFCC13	Dialect
1	-3.245.688	-3.186.747	-3.120.912	-32.146.894	-3.329.107	…….	4.187.491	West
2	-32.848.688	-32.132.829	-320.362.86	-3.292.678	-3.415.152	…….	4.022.270	West
3	-	-	-	-	-	…	-	We

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

	32.67 6.094	320.4 13.05	3.182 .289	32.32 2.820	33.68 4.574	… .	6.420 .176	st
4	- 32.18 6.078	- 3.149 .075	- 3.134 .935	- 31.91 8.075	- 3.329 .881	… … .	- 4.886 .838	We st
5	- 3.364 .437	- 3.273 .691	- 3.203 .315	- 32.59 8.596	- 3.320 .739	… … .	4.307 .632	We st
6	- 31.68 7.822	- 31.82 4.039	- 3.173 .723	- 32.72 9.593	- 3.377 .658	… … .	- 9.547 .366	We st
7	- 3.305 .284	- 3.285 .394	- 3.272 .659	- 3.361 .831	- 35.02 8.644	… … .	4.800 .604	We st
8	- 3.085 .055	- 2.969 .825	- 29.69 9.222	- 31.20 8.631	- 3.244 .687	… … .	5.396 .771	We st
9	- 3.262 .164	- 32.43 5.082	- 3.191 .192	- 3.271 .938	- 33.98 2.604	… … .	- 2.950 .396	We st
10	- 3.131 .350	- 2.987 .862	- 2.969 .718	- 299.7 96.63	- 30.63 6.218	… … .	7.292 .064	We st
…	…….	…….	…….	…….	…….	… … .	…….	… ….
100	- 3.255 .414	- 31.22 1.118	- 32.07 4.761	- 3.199 .827	- 33.01 9.434	… … .	- 32.01 3.347	Sho uth

a. *Pre-emphasis*

The purpose of pre-emphasis is to amplify the signal at high frequencies and reduce noise. With a constant filter. The calculation at the pre-emphasis stage is done as follows:

y[0] = 5.5348501
y[1] = y[1]-0.97(y[0]) = (-2.7390197)-0.97(5.5348501) = -8.1078242

The pre-emphasis process as shown in Figure 2 aims to enhance high-frequency components, which are vital for phoneme recognition and improve the performance of Random Forest and SVM in classifying Sundanese dialects. This process adds validity to the extraction and training of the model.

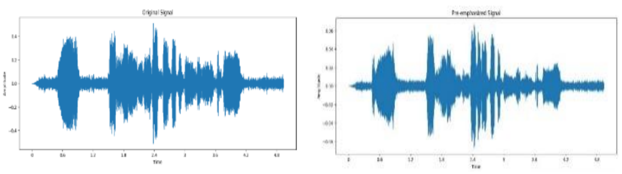

Fig. 2. Comparison Signal Before and After Pre-emphasis

b. Windowing

If the study of the function window is a hamming window, then the n-th function signal data window with the hamming window can be calculated using the n-th function signal data window calculation formula as follows:

w(0) = 0,54 – 0,46 (2π(0) 520)
= 0,54 – 0,46 cos(0) =0.08

After marking the nth window function obtained from the signal data, the windowing formula is used to calculate the windowing result x(n). The windowing result x(n) is obtained by the nth double sign signal frame signal (y(n)) with the window function w(n). The windowing calculation is done as follows:

x(0) =y(0) × w(0)
= (7,13790068) × 0,08 = 5.71032054

In Figure 3, it is shown that the use of Hamming window reduces the amplitude at the edges of the frame, while retaining it in the center, and helps reduce distortion and improve frequency analysis

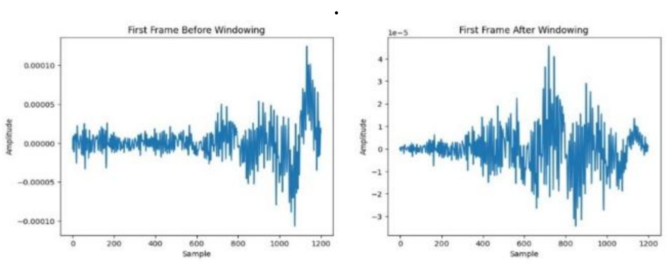

Fig. 3. Effect of windowing process on the first frame

c. Fast Fourier Transform (FFT)

The FFT process converts signals from the time domain to the frequency domain. With the number of samples N = 512, the calculation process in the Fast Fourier Transform process is as follows:

s(0) = ∑ 512−1 n = 0 x (n) e − j 2 πn (0) /512
s(0) = (5.71032054)e− j 2 π (0)(0) /512 + (8.19981570)e− j 2 π (1)(0) /512
+ (6.11131231)e− j 2 π (2)(0) /512 + (1.22343630)e− j 2 π (3)(0) /512
+ (7.83836882)e− j 2 π (4)(0) / 512 + (7.12290406)e− j 2 π (511)(0)/512
s(0) = 7.08910173

Figure 4 shows the amplitude spectrum of the first frame of the audio signal after going through the FFT process. There, the frequency (in Hertz) is displayed on the horizontal axis, while the vertical axis shows how much signal energy appears at each frequency.

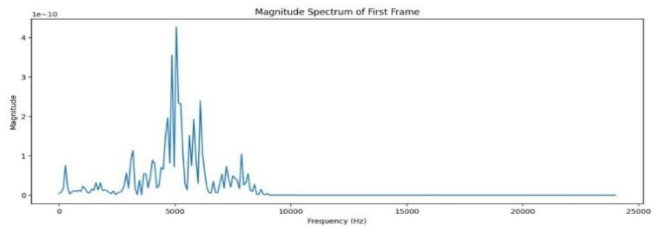

Fig. 4. First frame magnitude spectrum after FFT

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

d. *Mel Filter Bank*

The FFT energy spectrum is filtered using the Mel scale to approximate human hearing perception. The Mel Scale calculation is done as follows:
When zero frequency is entered into the Mel scale formula, the result is indeed zero too this is logical because there is no frequency that can be measured yet. But when we try to enter the number 5000 Hz into the same formula, the result is around 236.4. From there, we can see that in the context of this measurement, the lowest Mel value is 0, and the highest is around 236. So, this scale difference is what will later be used to adjust the perception of sound based on frequency.

e. *Discrete Cosine Transform (DCT)*

DCT can be used to compress frequency information and generate MFCC coefficients. For the calculation of the first coefficient, N is set to 40

$$c(o) = 2\sum_{n=0}^{40-1} \text{xncos}\frac{(2n+1)(0)}{2N}$$

=2[236.40458953cos(0) + (-234.26287304cos(0)) + (-208.0816647cos(0))
+...+(164.66021165)cos(0))] = -41.2797366
c(0) = (-41.2797366) x 1 /√4(40) = -3.24568825

The value of the first DCT coefficient is -3.24568825. The distribution of MFCC coefficients, which reflects the acoustic characteristics, is illustrated in Figure 5

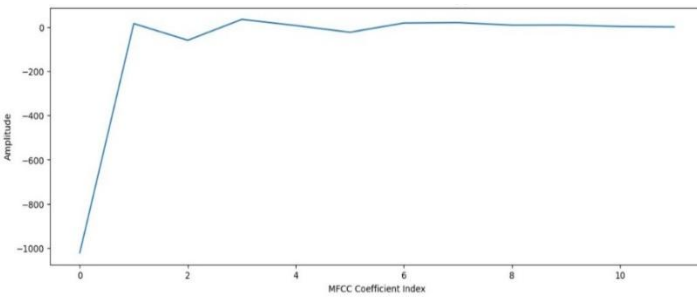

Fig. 5. Distribution of MFCC coefisients after discrete cosine transform

B. Classification Results Using Random Forest (RF) and Support Vector Machine (SVM) Algorithms

The performance of both models was evaluated using accuracy, precision, recall, and F1-score metrics. Additionally, the balance between true positive detection and false positive errors was analyzed using the ROC curve, while the overall quality of the models was assessed by measuring the Area Under the Curve (AUC).

a. *Analysis of Random Forest (RF) Algorithm Results*

The Random Forest model recorded an accuracy of 93.33%, with 28 out of 30 predictions correctly classified. For the West Sundanese dialect, the precision reached 0.89, recall 1.00, and F1-score 0.94. Meanwhile, for the Southern dialect, the precision was perfect at 1.00, recall 0.85, and F1-score 0.92. The ROC curve shows that this model is able to distinguish the two dialects well, as seen from the high true positive rate and low false positive rate at various decision thresholds.

Based on results displayed In Figure 6 Confusion Matrix for RF, we can understand how is this model succeed classifying voice data dialect in a way accurate . The RF model shows very high accuracy , namely 93.33%, which means that Of the 30 predictions made , 28 were correct . classified with Correct

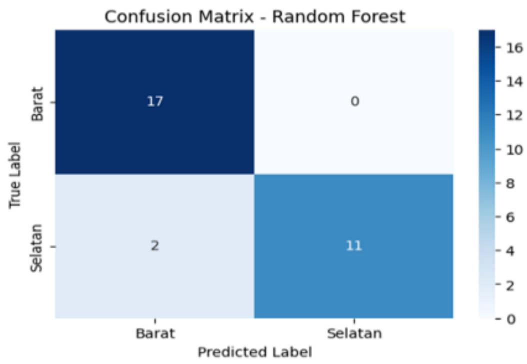

Fig. 6. Confusion matrix Random Forest

The following data is the result of the Random Forest method classification, shown in Figure 7.

Classification Report - Random Forest

	precision	recall	f1-score	support
Barat	0.89	1.00	0.94	17
Selatan	1.00	0.85	0.92	13
accuracy			0.93	30
macro avg	0.95	0.92	0.93	30
weighted avg	0.94	0.93	0.93	30

Fig. 7. Classification results with Random Forest

Figure 8 shows that Random Forest effectively differentiates West and South Sundanese dialects with high precision, recall, and F1-score. The model achieves 93% accuracy, with a precision of 0.89 and recall of 1.00 for the Western dialect, and precision of 1.00 with recall of 0.85 for the Southern dialect. The F1-scores (0.94 for West, 0.92 for South) indicate balanced performance. The ROC curve further confirms its strong classification ability..

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

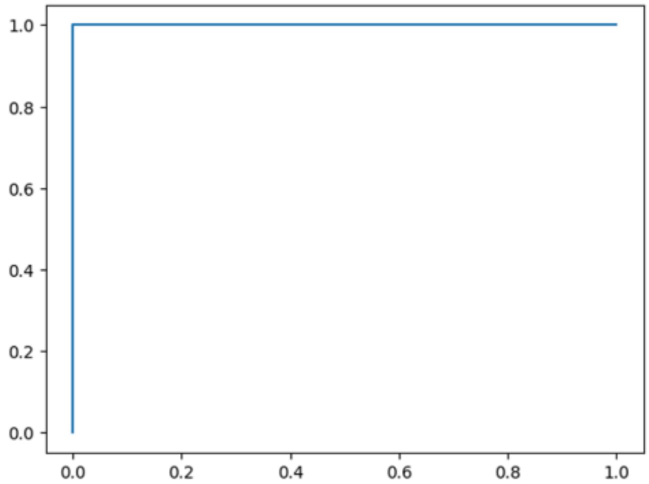

Fig. 8. ROC Curve for classification model with Random Forest

From the ROC curve above , it can be seen that the model has excellent performance , marked with almost curve approach corner left on graph . This is show that the model can differentiate second class with great accuracy on a wide range of decision threshold . The ROC line that reaches TPR value 1.0 without a significant increase in FPR indicates optimal performance , which means this model own level error very low prediction .

b. Analysis of Support Vector Machine (SVM) Algorithm Results

SVM managed to correctly recognize 14 samples from the Western dialect and 8 from the Southern. However, it still made some mistakes three samples that should have been negative were marked as positive, and five were missed entirely. This suggests that the model isn't very sensitive to the sound patterns of the Southern dialect. On the other hand, when applied to the Western dialect, its performance was fairly steady, with a precision of 0.74 and recall of 0.82. In contrast, the recall for the Southern dialect dropped to 0.62, which also caused its F1-score to fall to 0.67. The AUC came out to 0.81 decent, but not outstanding. Compared to Random Forest, SVM clearly fell short. RF delivered a higher accuracy of 93.33%, though it did require slightly more time and Recall to complete the process.

Different with RF, SVM shows more performance low in classification voice dialect , with accuracy overall 73.33%. Difference This possibility big due to the limitations of SVM in handle complex data distribution , especially when there is variation frequency non - linear sound . Figure 9 shows the Confusion Matrix for the applied Support Vector Machine (SVM) model. in classification West and South Sundanese dialects . Matrix This illustrate number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) in classification performed by the model.

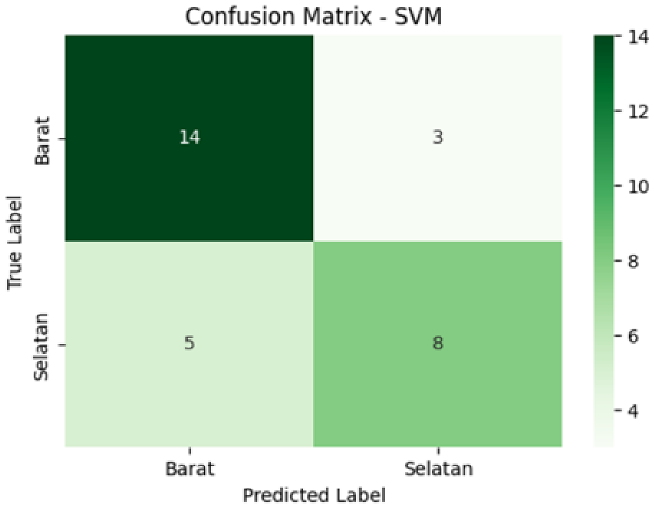

Fig. 9. Confusion matrix SVM

From the Confusion Matrix above , it can be seen that the SVM model has 14 True Positives (TP) for Western dialect and 8 True Positives (TN) for Southern dialect , showing correct prediction on both class . This model also recorded 3 False Positives (FP) for the West and 5 False Negatives (FN) for the South, indicating a number of error classification , especially in recognize Southern dialect . Although the SVM model has good performance in a way overall , the number of FNs is more high in the South class shows that the model is a bit difficulty in detect all example Southern dialect with appropriate .

The classification results of the Support Vector Machine (SVM) method are shown in Figure 10.

Classification Report - SVM

	precision	recall	f1-score	support
Barat	0.74	0.82	0.78	17
Selatan	0.73	0.62	0.67	13
accuracy			0.73	30
macro avg	0.73	0.72	0.72	30
weighted avg	0.73	0.73	0.73	30

Fig. 10. Classification results with Support Vector Machine

The SVM model performs better on the Western dialect (precision 0.74, recall 0.82, F1-score 0.78) than the Southern dialect (recall 0.62, F1-score 0.67), with an overall accuracy of 73%. Figure 11 presents the ROC Curve, illustrating the model's ability to distinguish both dialects based on TPR and FPR.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

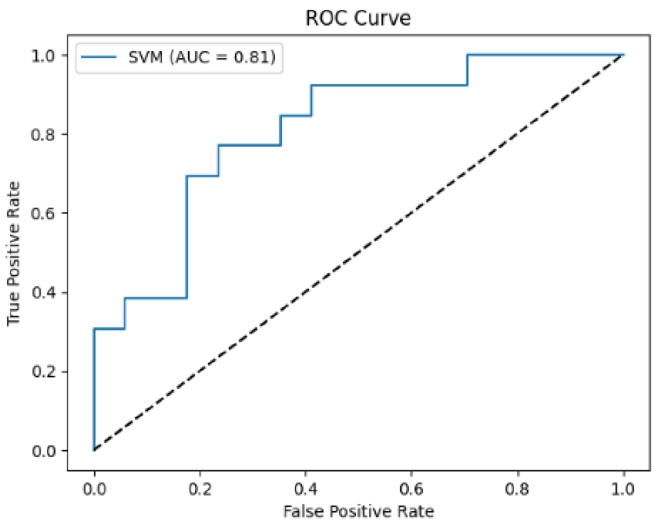

Fig. 11. ROC Curve for classification model with SVM

The ROC Curve in Figure 13 shows that the SVM model has an AUC of 0.81, indicating good classification performance but with room for improvement in distinguishing West and South Sundanese dialects. Table 3 compares RF and SVM based on processing time, accuracy, and memory usage.

TABLE II. COMPARISON RANDOM FOREST AND SVM ALGORITHMS

Algorithm	Processing Time	Accuracy	Memory
Random Forest (RF)	2.3 seconds	93.33%	520 KB
SVM	1.8 seconds	73.33%	480 KB

Table 2 shows the comparison items consists of from Process Time namely the duration required for each algorithm For carry out the classification process , accuracy is level accuracy prediction from each algorithm against test data, memory : Memory is required by every algorithm during the classification process . From the table this , looks that Random Forest has more precision tall compared to SVM, although need little processing time and memory more big.

The performance gap between RF and SVM shows how ensemble techniques can better generalize on constrained, high-dimensional voice data.

Misclassifications identified in the confusion matrix indicate that specific dialectal characteristics may intersect, presenting a problem for boundary-based classifiers such as SVM. These overlaps illustrate the practical challenge of differentiating regional dialects that possess common phonetic origins. Additional enhancements could be realized by integrating a broader range of audio properties or employing more sophisticated models with temporal sensitivity.

IV. CONCLUSION

The results of this study indicate that both Random Forest (RF) and Support Vector Machine (SVM) are effective in classifying West and South Sundanese dialects using MFCC features since RF consistently outperforms SVM across accuracy, precision, recall, and F1-score. RF was particularly consistent in identifying the Southern dialect, indicating its superiority in relation to possible feature change.

Nevertheless, this study has certain restrictions. The small size of the dataset can restrict the model's generalizability. Moreover, the absence of data augmentation or noise-handling methods could affect performance in real-world scenarios where audio quality varies.

Future research is fascinating to explore using deep learning models as convolutional neural networks (CNNs) or recurrent neural networks (RNNs). These models offer more feature learning capability and better scalability for larger and more complex datasets. They can also adapt to minor acoustic patterns that standard machine learning models overlook, hence perhaps enhancing accurate and generalizable dialect detection techniques.

REFERENCES

[1] M. Azhar and H. F. Pardede, “Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest,” J. MEDIA Inform. BUDIDARMA, vol. 5, no. 2, p. 439, Apr. 2021, doi: 10.30865/mib.v5i2.2754.

[2] A. M. Afida, “KLASIFIKASI JENIS BURUNG BERDASARKAN SUARA MENGGUNAKAN ALGORITME SUPPORT VECTOR MACHINE,” MATHunesa J. Ilm. Mat., vol. 8, no. 1, pp. 1–6, Jan. 2020, doi: 10.26740/mathunesa.v8n1.p1-6.

[3] Akhiril Anwar Harahap, R. Novita, T. K. Ahsyar, and Z. Zarnelly, “Classification of Beef and Pork with Deep Learning Approach,” J. Sist. Cerdas, vol. 7, no. 1, pp. 55–65, Apr. 2024, doi: 10.37396/jsc.v7i1.393.

[4] A. Alam, S. Urooj, and A. Q. Ansari, “Design and Development of a Non-Contact ECG-Based Human Emotion Recognition System Using SVM and RF Classifiers,” Diagnostics, vol. 13, no. 12, p. 2097, Jun. 2023, doi: 10.3390/diagnostics13122097.

[5] A. Aljohani, N. Alharbe, R. E. Al Mamlook, and M. M. Khayyat, “A hybrid combination of CNN Attention with optimized random forest with grey wolf optimizer to discriminate between Arabic hateful, abusive tweets,” J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 2, p. 101961, Feb. 2024, doi: 10.1016/j.jksuci.2024.101961.

[6] M. A. As Sarofi, I. Irhamah, and A. Mukarromah, “Identifikasi Genre Musik dengan Menggunakan Metode Random Forest,” J. Sains Dan Seni ITS, vol. 9, no. 1, pp. D79–D86, Jun. 2020, doi: 10.12962/j23373520.v9i1.51311.

[7] C. Avci, M. Budak, N. Yağmur, and F. Balçik, “Comparison between random forest and support vector machine algorithms for LULC classification,” Int. J. Eng. Geosci., vol. 8, no. 1, pp. 1–10, Feb. 2023, doi: 10.26833/ijeg.987605.

[8] C. Doğdu, T. Kessler, D. Schneider, M. Shadaydeh, and S. R. Schweinberger, “A Comparison of Machine Learning Algorithms and Feature Sets for Automatic Vocal Emotion Recognition in Speech,” Sensors, vol. 22, no. 19, p. 7561, Oct. 2022, doi: 10.3390/s22197561.

[9] G. Drougkas, E. Bakker, and M. Spruit, “Multimodal Machine Learning for Language and Speech Markers Identification in Mental Health,” Sep. 27, 2024, In Review. doi: 10.21203/rs.3.rs-4925232/v1.

[10] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” J. Inform. Upgris, vol. 7, no. 1, Jun. 2021, doi: 10.26877/jiu.v7i1.7099.

[11] Q. Du, J. Shen, P. Wen, and X. Chen, “Parkinson’s Disease Detection by Using Machine Learning Method based on Local Classification on Class Boundary,” Discov. Appl. Sci., vol. 6, no. 11, p. 576, Oct. 2024, doi: 10.1007/s42452-024-06295-1.

[12] U. Farooq, K. K. S. Reddy, K. S. Shishira, M. G. Jayanthi, and P. Kannadaguli, “Comparing Hindustani Music Raga Prediction Systems using DL and ML Models,” in 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary

Applications (ICETCS), Bengaluru, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICETCS61022.2024.10543647.

[13] S. Garg and B. Raghavan, “Comparison of machine learning algorithms for the classification of spinal cord tumor,” Ir. J. Med. Sci. 1971 -, vol. 193, no. 2, pp. 571–575, Apr. 2024, doi: 10.1007/s11845-023-03487-3.

[14] R. B. Handoko and S. Suyanto, “Klasifikasi Gender Berdasarkan Suara Menggunakan Support Vector Machine,” Indones. J. Comput. Indo-JC, vol. 4, no. 1, p. 9, Mar. 2019, doi: 10.21108/INDOJC.2019.4.1.244.

[15] O. Peña-Cáceres, H. Silva-Marchan, M. Albert, and M. Gil, “Recognition of Human Actions through Speech or Voice Using Machine Learning Techniques,” Comput. Mater. Contin., vol. 77, no. 2, pp. 1873–1891, 2023, doi: 10.32604/cmc.2023.043176.

[16] S.-M. Jeong, Y.-D. Song, C.-L. Seok, J.-Y. Lee, E. C. Lee, and H.-J. Kim, “Machine learning-based classification of Parkinson’s disease using acoustic features: Insights from multilingual speech tasks,” Comput. Biol. Med., vol. 182, p. 109078, Nov. 2024, doi: 10.1016/j.compbiomed.2024.109078.

[17] M. Kasahun and A. Legesse, “Machine learning for urban land use/ cover mapping: Comparison of artificial neural network, random forest and support vector machine, a case study of Dilla town,” Heliyon, vol. 10, no. 20, p. e39146, Oct. 2024, doi: 10.1016/j.heliyon.2024.e39146.

[18] S. Madanian et al., “Speech emotion recognition using machine learning — A systematic review,” Intell. Syst. Appl., vol. 20, p. 200266, Nov. 2023, doi: 10.1016/j.iswa.2023.200266.

[19] G. H. Mohmad Dar and R. Delhibabu, “Speech Databases, Speech Features, and Classifiers in Speech Emotion Recognition: A Review,” IEEE Access, vol. 12, pp. 151122–151152, 2024, doi: 10.1109/ACCESS.2024.3476960.

[20] A. Omar and T. Abd El-Hafeez, “Quantum computing and machine learning for Arabic language sentiment classification in social media,” Sci. Rep., vol. 13, no. 1, p. 17305, Oct. 2023, doi: 10.1038/s41598-023-44113-7.

[21] D. Subhash, J. L. G., P. B., and V. Ravi, “A robust accent classification system based on variational mode decomposition,” Eng. Appl. Artif. Intell., vol. 139, p. 109512, Jan. 2025, doi: 10.1016/j.engappai.2024.109512.

[22] M. Ur Rehman, A. Shafique, Q.-U.-A. Azhar, S. S. Jamal, Y. Gheraibia, and A. B. Usman, “Voice disorder detection using machine learning algorithms: An application in speech and language pathology,” Eng. Appl. Artif. Intell., vol. 133, p. 108047, Jul. 2024, doi: 10.1016/j.engappai.2024.108047.

[23] N. Widjiyati, “Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval,” J. Janitra Inform. Dan Sist. Inf., vol. 1, no. 1, pp. 1–7, Apr. 2021, doi: 10.25008/janitra.v1i1.118.

[24] N. Zhantileuov and S. Ospanov, “A Comparative Study of Supervised Machine Learning and Deep Learning Techniques with Feature Selection Methods for Classifying Parkinson’s Disease Based on Speech Impairments,” in 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan: IEEE, May 2024, pp. 124–129. doi: 10.1109/SIST61555.2024.10629274.

[25] E. Raczko and B. Zagajewski, “Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images,” Eur. J. Remote Sens., vol. 50, no. 1, pp. 144–154, Jan. 2017, doi: 10.1080/22797254.2017.1299557.

[26] E. Renata and M. Ayub, “Penerapan Metode Random forest untuk Analisis Risiko pada dataset Peer to peer lending,” J. Tek. Inform. Dan Sist. Inf., vol. 6, no. 3, Dec. 2020, doi: 10.28932/jutisi.v6i3.2890.

[27] A. B. Dina, R. Sarno, R. N. E. Anggraini, A. T. Haryono, and A. F. Septiyanto, “Comparison of Oversampling Techniques in Prediction Judicial Decisions of Divorce Trials in Family Courts,” in 2024 International Conference on Information Technology Research and Innovation (ICITRI), Jakarta, Indonesia: IEEE, Sep. 2024, pp. 13–18. doi: 10.1109/ICITRI62858.2024.10699016.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025