RESEARCH ARTICLE OPEN ACCESS

Manuscript received November 1, 2023; revised December 30, 2023; accepted December 30, 2023; date of publication January 20, 2024
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v6i1.350

How to cite: Putri Agustina Riadi, Mohammad Reza Faisal, Dwi Kartini, Radityo Adi Nugroho, Dodon Turianto Nugrahadi, and Dike Bayu Magfira, A Comparative Study of Machine Learning Methods for Baby Cry Detection Using MFCC Features, Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 1, pp. 73-83, January 2024.

A Comparative Study of Machine Learning Methods for Baby Cry Detection Using MFCC Features

Putri Agustina Riadi[1], Mohammad Reza Faisal[1], Dwi Kartini[1], Radityo Adi Nugroho[1], Dodon Turianto Nugrahadi[1], Dike Bayu Magfira[2]

[1]Computer Science Department, Lambung Mangkurat University, Banjarbaru, South Kalimantan, Indonesia
[2]Information System Department, Universitas Nahdlatul Ulama Surabaya, Surabaya, Jawa Timur, Indonesia

Corresponding author: Mohammad Reza Faisal (e-mail: reza.faisal@ulm.ac.id)

ABSTRACT Baby crying is one of the main ways babies communicate with their parents to convey their needs and emotions. While the act of baby crying can yield crucial insights into a baby's needs and emotions, there is a dearth of research explicitly investigating the influence of the audio range within a baby cry on research outcomes. The core research problem is the lack of research on the influence of audio range on baby cry classification on machine learning. This research aims to compare the effect of the audio length of a baby's cry in MFCC feature extraction and several machine learning algorithms on the performance of baby emotion detection. The contribution enriches an understanding of classification and feature selection applications in audio datasets, particularly in baby cry audio. The utilized dataset, donate-a-cry-corpus, encompasses five distinct data classes and possesses seven seconds. The employed methodology consists of the spectrogram technique, cross-validation for data partitioning, MFCC feature extraction with 10, 20, and 30 coefficients, and machine learning models including Support Vector Machine, Random Forest, and Naïve Bayes. The findings of this study reveal that the Random Forest model achieved an accuracy of 0.844 and an F1 score of 0.773 when 10 MFCC coefficients were utilized and the optimal audio range was set at six seconds. Furthermore, the Support Vector Machine model with an RBF kernel yielded an accuracy of 0.836 and an F1 score of 0.761. In contrast, the Naïve Bayes model achieved an accuracy of 0.538 and an F1 score of 0.539. Notably, no discernible differences were observed when evaluating the Support Vector Machine and Naïve Bayes methods across the 1-7 second trial. The implication of this research is to establish a foundation for advancing premature illness identification techniques grounded in the vocalizations of baby, thereby facilitating swifter diagnostic processes for pediatric practitioners.

INDEX TERMS Baby cry detection, Spectrogram, MFCC, machine learning.

I. INTRODUCTION

The vocalization of distress through crying is the primary mode of communication employed by baby, as it allows them to express their needs and emotions effectively. About 130 million babies are born globally each year [1]. Understanding a baby's cry is crucial in providing adequate care, enabling parents to accurately respond to the baby's needs. However, understanding the meaning of crying is difficult for many people, especially new parents. Although a baby cry can provide important clues about the baby's needs and emotions, few studies have specifically examined the influence of the audio range and best method in a baby cry on research results. The problem needs to be investigated because it can affect a baby's health. Machine Learning (ML) algorithms are being explored and applied for numerous tasks in clinical workflows ranging from disease prognosis, diagnosis, medical treatment, defining a care plan for the patient, and many more [2]. Machine learning, a brand of computational science, is extensively used in this part, with the classification task as the major approach [3].

In the initial stages of baby cry research [4], the K-means clustering method and Gaussian mixture models were employed, yielding an accuracy rate of 81.27% when utilizing the donate-a-cry corpus dataset. Subsequently, further research on baby cry analysis was conducted by [5], who employed the CNN algorithm with a smaller training dataset and achieved an accuracy rate of 72% using online sound libraries of baby cries without segmenting the audio into seconds.

In this study, the baby cry audio files undergo a preprocessing stage to identify significant features. The Mel Frequency Cepstrum Coefficient (MFCC) is a feature extraction method utilized in audio research. MFCC enables the processing of audio variations by converting sound signals into MFCC coefficients represented as a vector sequence, which can then be employed for research purposes. In [6], MFCC yielded a 96% accuracy rate. Cross-validation is a data resampling method that aims to assess the generalization ability of predictive models and prevent overfitting [7]. Specifically, the study employed K-Fold cross-validation to classify baby cries, utilizing a 5-fold cross-validation approach. [8]

In the context of baby cry datasets, the accuracy of machine learning methods is influenced. One such case involves the usage of SVM, which yielded an F1 score of 10.3% [9]. Another case involves the application of SVM on the baby Chillanto database with 5-fold cross-validation, resulting in an accuracy rate of 90% [8]. The study by [10] utilized SVM with an RBF kernel, achieving an accuracy rate of 0.560. In addition to SVM, the study also employed the Naïve Bayes method, which demonstrated good accuracy in classifying gender datasets with a rate of 87% [6]. Furthermore, SVM was utilized in predicting health issues, with the study combining SVM with XG Boost and achieving an accuracy rate of 85.71% [11]. SVM had high accuracy in the study [12], which resulted in 96% accuracy.

Random Forest has been utilized in previous research [12]–[14] and has demonstrated the highest accuracy ranging from 62% to 80%. Following a comprehensive comparison of various machine learning methods employed in previous studies, it is deemed advantageous to adopt these three methods and utilize MFCC feature extraction in the present study due to their ability to yield favourable accuracy outcomes.

Distinct from preceding research endeavours, this study diverges because it explores a wider range of duration variations in the classification of baby cry audio. The novelty of this research lies in utilizing different coefficient MFCC values (10, 20, 30) on different audio lengths of a baby's cry (1 – 7 seconds) by using three classification algorithms, namely Support Vector Machine, Random Forest, and Naïve Bayes.

This research aims to determine the prediction performance of several classification models that have been built. The models are built based on different combinations of MFCC coefficient values, lengths of a baby's cry, and classification algorithms. The performance obtained from each model will provide knowledge of which combination configuration can detect emotions based on a baby's cry. So this model can then be applied to create an application for detecting baby emotions.

The results of this research are expected to provide contributions such as:

a. It provides a better understanding of classification performance based on the audio length and the number of feature extraction coefficients on audio datasets, specifically for audio recordings of baby crying.
b. It offers insight into the most effective algorithm for classifying audio signals produced by baby crying.
c. It has the potential to be implemented as an intelligent application aimed at assisting parents in identifying the emotions and needs of their babies through the analysis of different cry types.
d. The results of this study further enrich the existing body of knowledge on audio classification, with a specific focus on the domain of baby crying audio.

II. MATERIAL AND METHODS

The research flow of this research can be seen in FIGURE 1 which consist of data collection, pre-processing, K-fold cross validation, and classification.

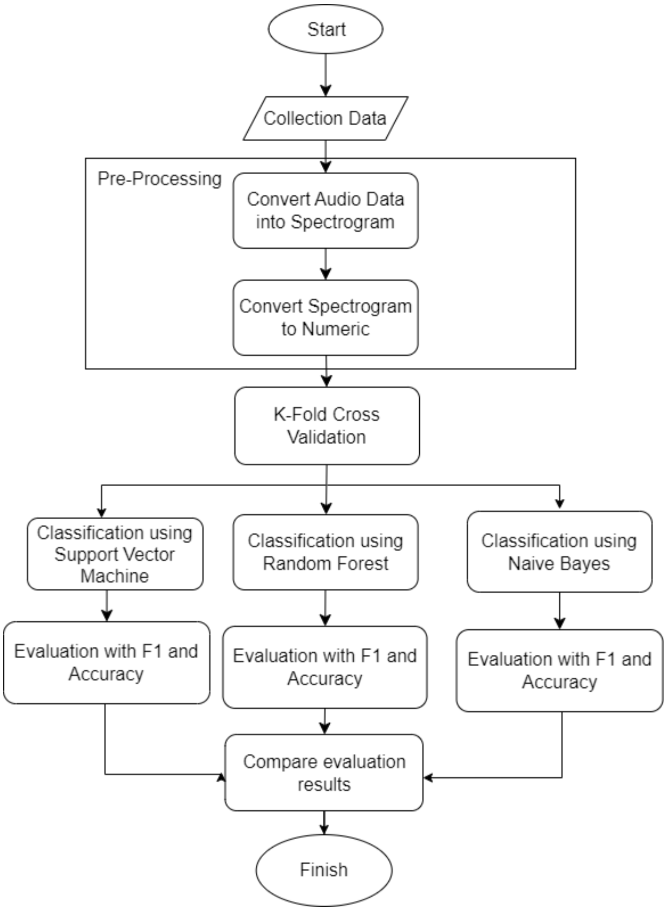
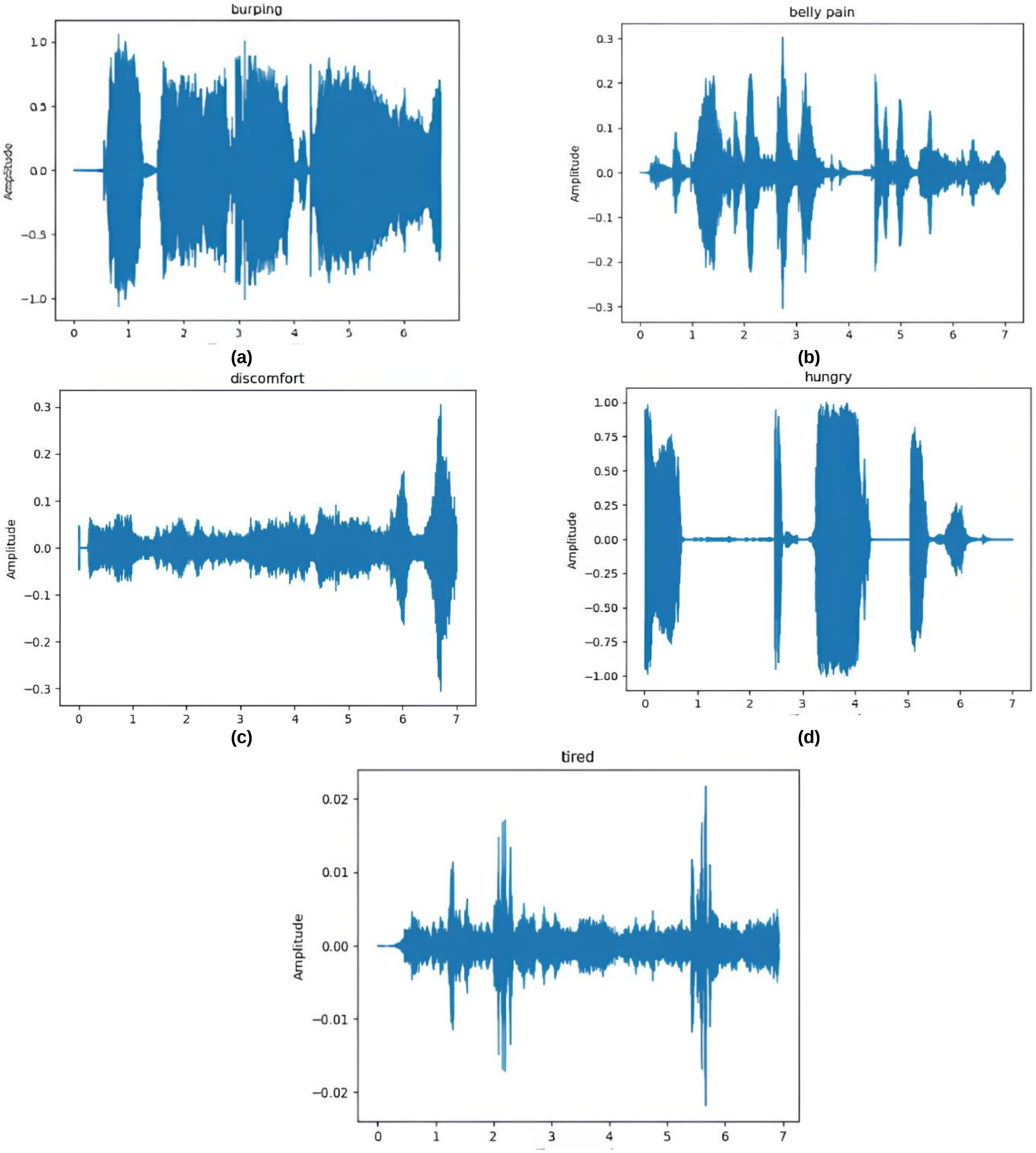

Start
Collection Data
Pre-Processing
Convert Audio Data into Spectrogram
Convert Spectrogram to Numeric
K-Fold Cross Validation
Classification using Support Vector Machine
Classification using Random Forest
Classification using Naive Bayes
Evaluation with F1 and Accuracy
Evaluation with F1 and Accuracy
Evaluation with F1 and Accuracy
Compare evaluation results
Finish

FIGURE 1 **Research Flow of SVM, Random Forest, and Naïve Bayes Classification Models**



(a) (b)

(c) (d)

(e)

FIGURE 2. The sound wave for baby a) crying burping, b) cry bell-pain, c)cry discomfort, d)cry hungry, and e) cry tired.

A. DATASET

Donate-a-cry-corpus is the audio dataset of baby crying recordings that have been used in previous research [12], [15]–[17]. This dataset has five class labels:

- Class label 0 is belly pain. This label consists of 16 audios.
- Class label 1 is burping and consists of 8 audios.
- Class label 2 is discomfort and consists of 27 audios.
- Class label 3 is hungry and consists of 383 audios.
- Class label 4 is tired and consists of 24 audios.

The total recording in this dataset is 458 recorded audios that are 7 seconds long and formatted as WAV. The sound wave graph for the burping label can be seen in FIGURE 2. TABLE 1 shows details of the dataset that consists of two columns. The first column is the audio file name, and the second is the class label.

TABLE 1
Detail Dataset

No	File	Emotion
1	hu_e4051e62-d21d-4bb8-a235-fd7e859ad787-1430740613780-1.7-f-72-hu.wav	hu
2	hu_045C5483-69E1-4BEC-B1D8-9286D174B9B2-1430102996-1.0-m-04-hu.wav	hu
….	….	….
457	hu_6d922623-8424-4625-be4c-4964e0c9e25c-1434898452774-1.7-m-04-hu.wav	hu
458	hu_B327333E-2DE2-4833-A75A-4C576208BED3-1430087456-1.0-f-48-hu.wav	hu

B. PREPROCESSING

Preprocessing is improving and facilitating the quality of raw data so that it can later be used in further steps [18]. Data reduction involves removing unimportant data and selecting essential data [19]. In this phase, the work process is divided into two parts, specifically transforming raw audio data into two-dimensional data known as a spectrogram. This section is to convert audio to dataset. The preprocessing that will be conducted in this study entails converting audio data into visual data in the form of spectrograms. Spectrograms are commonly employed in the domains of acoustic science, audio engineering, and related disciplines. Consequently, the transformed data can be observed in FIGURE 3.

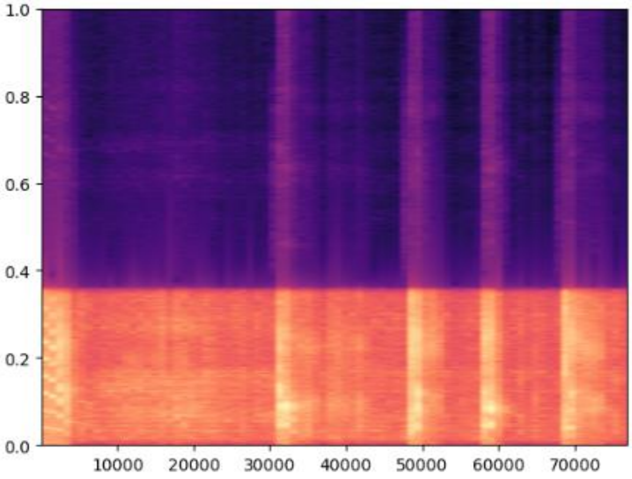


FIGURE 3. Spectrogram

The initial audio data encompasses 7 seconds, and within this temporal span, experiments will be conducted at intervals of 1 second, 2 seconds, 3 seconds, 4 seconds, 5 seconds, 6 seconds, and 7 seconds. These experiments aim to ascertain whether any disparities in the research findings arise. Subsequently, the data will undergo spectrogram processing, followed by the extraction of features. Specifically, six features will be utilized for extraction, namely Chroma Short Time Fourier Transform, Root Mean Square, Spectral Bandwidth, Spectral Rolloff, Zero Crossing Rate, and Mel Frequency Cepstral Coefficient. This feature extraction is used because research [10] obtained good results with this feature.

The feature extraction process on the 7-second audio can be observed in the following TABLE 2. In this table, all feature that are used in preprocessing are shown. The remaining information is encapsulated within the table, referring to the first to tenth Mel Frequency Cepstral Coefficients (MFCC 1-10). Feature extraction with 6-second audio can be seen in TABLE 3. In this table, there is all feature that is used in preprocessing. The remaining information is encapsulated within the table, referring to MFCC 1-10. There is a different result in Table 2 because seconds affect feature extraction. Feature extraction with 5-second audio can be seen in TABLE 4. There is a different result in TABLE 3 because seconds affect feature extraction. Feature extraction was also conducted for 1-4 seconds, yielding favorable outcomes. Consequently, this process can be further pursued to extract MFCC.

TABLE 2
Feature Extraction Results For 7 Seconds

No	Croma_stft	RMSE	Spectral_centroid	…	Mfcc10
1	0.313	0.106	1224.184	…	-7.494
2	0.273	0.165	1038.469	…	-4.836
…	…	…	…	…	…
457	0.278	0.0516	729.203	…	-5.456
458	0.442	0.035	1466.236	…	2.444

TABLE 3
Feature Extraction Results For 6 Seconds

No	Croma_stft	RMSE	Spectral_centroid	…	Mfcc10
1	0.329	0.094	1234.265	…	-7.887
2	0.296	0.173	1072.189	…	-4.289
…	…	…	…	…	…
457	0.277	0.053	757.355	…	-4.985
458	0.464	0.034	1496.897	…	2.053

TABLE 4
Feature Extraction Results For 6 Seconds

No	Croma_stft	RMSE	Spectral_centroid	…	Mfcc10
1	0.329	0.091	1264.557	…	-9.353
2	0.273	0.204	1064.721	…	-3.532
…	…	…	…	…	…
457	0.282	0.052	748.069	…	-6.446
458	0.471	0.026	1461.789	…	0.825

C. MFCC FEATURE EXTRACTION

Mel Frequency Cepstral Coefficients (MFCC) is also often used in speech recognition [1], [9], [20]. The utilization of MFCC for feature extraction stems from its ability to mimic the functionality of the human auditory system, which involves logarithmically filtering sound signals. MFCC refers to the capability of algorithms to generate minimal data while retaining crucial information inherent in an audio signal. Mel spectrogram converts audio signals into a spectrogram with a mel scale [20]. Feature extraction is obtaining characteristics to characterize an image [21].

$$MFCC(n,m) = \frac{1}{N}\sum_{k=0}^{N-1} log\left(\sum_{k=0}^{N-1} X(k,n)\, e^{-j2\pi km/N}\, H_{m(k)}\right. \qquad (1)$$

In equation (1) [22], n represents audio frame, m is a cepstral coefficient, N dedicates to the number of frames used in the fast Fourier transform, $X\ (k,n)$ is the FFT value of frame n at frequency k, and $\mathrm{H_m}$ (k) is filterbank Mel -m at frequency k.

Features of MFCC coefficients 10, 20, and 30 will be extracted from the acquired dataset. Subsequently, the number of features will be augmented by an additional six extraction features. In the case of using a coefficient of 10, the number of features will amount to 16. Similarly, if a coefficient of 20 is employed, the number of features will increase to 26. Lastly, if a coefficient of 30 is utilized, the number of features will reach 36. These outcomes concern the original audio duration of the dataset.

TABLE 5 shows the results for 10 MFCC. The table displays the process of extracting features from a 7-second audio clip. This process involves extracting 6 features and the Mel-frequency cepstral coefficients (MFCC) 1-10. As a result, the initial set of features is expanded to a total of 16 features. Another result for 20 MFCC is presented in TABLE 6. The table displays the process of extracting features from a 7-second audio clip. This process involves extracting 6 features and the Mel-frequency cepstral coefficients (MFCC) 1-20. As a result, the initial set of features is expanded to a total of 26 features. Also, 30 MFCC are presented in TABLE 7. The table displays the process of extracting features from a 7-second audio clip. This process involves extracting 6 features and the Mel-frequency cepstral coefficients (MFCC) 1-30. As a result, the initial set of features is expanded to 36 features.

TABLE 5
Result for 10 MFCC Feature Extraction

No	Croma_stft	RMSE	Spectral_centroid	…	MFCC10
1	0.313	0.106	1224.184	…	-7.494
2	0.273	0.165	1038.469	…	-4.837
…	…	…	…	…	…
457	0.279	0.051	729.203	…	-5.456
458	0.442	0.034	1466.236	…	2.444

TABLE 6
Result For 20 MFCC Feature Extraction

No	Croma_stft	RMSE	Spectral_centroid	…	MFCC20
1	0.313	0.106	1224.184	…	-4.337
2	0.273	0.165	1038.469	…	-4.219
…	…	…	…	…	…
457	0.279	0.052	729.203	…	-6.421
458	0.442	0.035	1466.236	…	-9.497

TABLE 7
Result for 30 MFCC Feature Extraction

No	Croma_stft	RMSE	Spectral_centroid	…	MFCC30
1	0.313	0.106	1224.184	…	-3.656
2	0.273	0.165	1038.469	…	-2.193
…	…	…	…	…	…
457	0.279	0.052	729.203	…	-3.568
458	0.442	0.035	1466.236	…	-0.398

D. 10 K-FOLD CROSS-VALIDATION

Cross-validation is a statistical method for evaluating the performance of an algorithm that has been designed. The cross-validation capability is that it can divide training and testing data [23]. Cross-validation is a computational method requiring information partitioning through subsets [24]. Cross-validation is also resampling data to prevent overfitting [7]. One part is used to validate the model, and the rest to train the classifier [25]

This stage divides the dataset into training and test data using cross-validation with a value of k = 10. The data will be divided into ten subsets with the same class number [10].

E. SUPPORT VECTOR MACHINE CLASSIFICATION

Support Vector Machine is an algorithm for predicting and classifying linear functions in high-dimensional space. Support vector machines are also used for feature training and testing [6]. SVM is also used for regression. SVM is one of the machine learning methods that is easy to implement [21]. SVM is a binary classification model [26], [6].

The main focus of this algorithm is to build an optimal hyperplane (separator) to separate two different classes of data. An SVM display depicts instances as points in space, strategically arranged to ensure that the models of distinct categories are segregated by a meaningful, maximally wide gap [27]. The mathematical underpinnings of Support Vector Machines (SVMs) are deeply rooted in the construction of a hyperplane, which is defined by equation (2) [27].

$$\omega \cdot x + b = 0 \qquad (2)$$

$$yi \cdot (xi + b) \geq 1 \qquad (3)$$

where the weight vector ω is orthogonal to the hyperplane, while b represents the bias term that determines the hyperplane's offset from the origin. From a mathematical standpoint, SVMs address an optimization problem by minimizing ½ |(|ω|)|^2. Subject to the constraints outlined in equation (3) for all data points, with yi denoting the class labels, and X_i is the represent list of x.

This particular formulation guarantees not only accurate classification but also a significantly wide margin, thereby rendering SVMs a reliable and versatile choice for various machine-learning endeavors. SVM can be used in formula (4) in linear cases.

$$K\ (x_{i,}x_j) = x.y \tag{4}$$

where K is the kernel used on SVM, x and y are points in the data that form a vector representing values in the classification. Using kernels SVM for data classification needs that cannot be solved linearly. One of these kernels is RBF. The following formula is the RBF kernel SVM equation [22].

$$K\left(x_{i,}x_j\right) = \exp\left(-\frac{||X_i - X_j||^2}{2\sigma^2}\right) \tag{5}$$

where K is dedicated to the kernel used on SVM. *x* and *y* are points in the data form a vector representing values in the classification, and σ is the parameters used in the RBF kernel. SVM is divided into two kernels, namely RBF and linear. This linear kernel is a well-known and popular SVM kernel. The RBF kernel is a kernel concept that aims to classify data that cannot be separated linearly. This research was also used in [12].

F. RANDOM FOREST CLASSIFICATION

The Random Forest (RF) algorithm is an ensemble method in machine learning. It builds multiple decision trees and combines their outputs for more accurate predictions [28], [25], [12], [29]. Each tree is constructed using a random subset of the data, and the final prediction is determined by a vote from all the trees [14], [24]. This approach enhances accuracy, reduces overfitting, and works well for both classification and regression tasks.

The Random Forest algorithm's formula is used in equation (6) [22].

$$Gini\ (S) = 1 - \sum pi2\ k\ i = 1 \tag{6}$$

where *pi* is the probability of S belonging to class i, and k is the dataset's number of classes or categories. Pi represents the proportion of the dataset belonging to class or category i. This algorithm proceeds with the following steps [12]:

1. Select random samples from the database.
2. Construct a decision tree for each sample. Obtain the prediction from each decision tree.
3. Count the frequency of results for each class
4. Select the most frequent result as the final prediction

G. NAÏVE BAYES CLASSIFICATION

The Naive Bayes algorithm is a probabilistic classification technique based on Bayes' theorem. It assumes that the features used for classification are independent, which might be an oversimplification in real-world scenarios [6], [30]. The algorithm calculates the probability of a data point belonging to a certain class given its feature values. During training, it learns the probabilities from the data. In prediction, it multiplies the probabilities of individual features for each class and selects the class with the highest probability as the final prediction. Despite its simplicity and the "naive" assumption of independence, Naive Bayes often performs well in text classification and spam filtering tasks [24]. The advantage of using NB is that it only requires training data that is not large to determine the estimated parameters needed in the classification process [6].

$$(C|F_{1,} \dots , F_n) = \frac{p\ (C)p(F_{1,\dots,}F_n|C)}{p(F_{1,\dots,}F_n)} \tag{7}$$

where $p(C|F_{1,} \dots , F_n)$ is the posterior probability, p (C) is the probability of class C. $p\left(F_{1,} \dots , F_n\,\middle|C\right)$ is the probability likelihood, and $p(F_{1,} \dots , F_n)$ prior probability of the instance $(F_{1,} \dots , F_n)$.

III. RESULTS

This section presents the performance of models for detecting baby crying, utilizing Support Vector Machine, Random Forest, and Naïve Bayes algorithms. The duration of the audio samples of baby cries falls within the range of 1 to 7 seconds, while the number of MFCC coefficients exhibits variability.

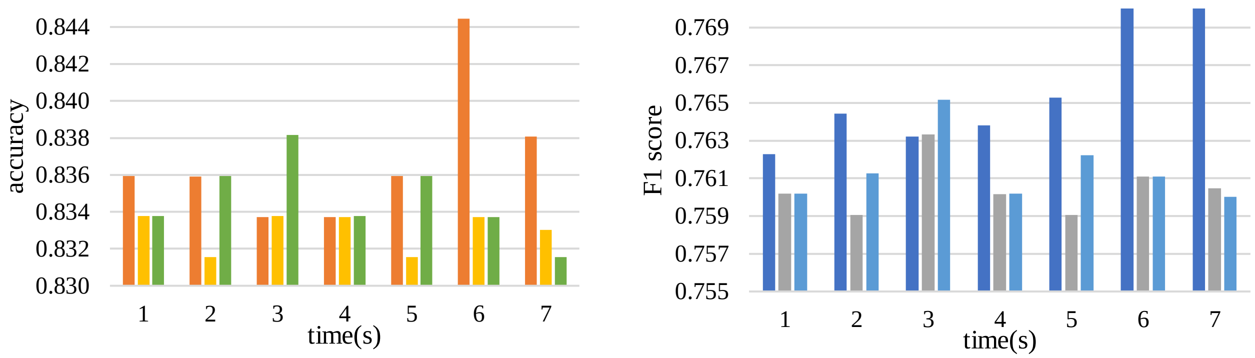
A. PERFORMANCE OF RANDOM FOREST

The SVM classification model is constructed using the identical parameters as the study [10], which are as follows: n_estimators = 100, random_state = 42, n_splits = 10, and shuffle = True. The performance of this classification model can be observed in FIGURE 4. Model performance, evaluated by accuracy measurements, can be found in section (a), while section (b) presents the F1 score measurements. The performance value, utilizing 10 MFCC coefficients, is visualized in the first bar, and the subsequent bar demonstrates the performance value when the number of the MFCC coefficients is 20 and 30.

The model that uses feature extraction data with 10 MFCC coefficients produces the highest accuracy of 0.836 and F1 score of 0.762. That highest performance value is obtained when using 6 seconds of audio. The best performance of the model built with data featuring 20 MFCC coefficients is that the accuracy value is 0.763, and the F1 score is 0.834. That result was obtained using audio with a length of 3 seconds. Moreover, a model using data created with 30 MFCC coefficients performed best when using 3 seconds of audio. That model's performance has an accuracy of 0.838 and an F1 score of 0.765.

B. PERFORMANCE OF SUPPORT VECTOR MACHINE

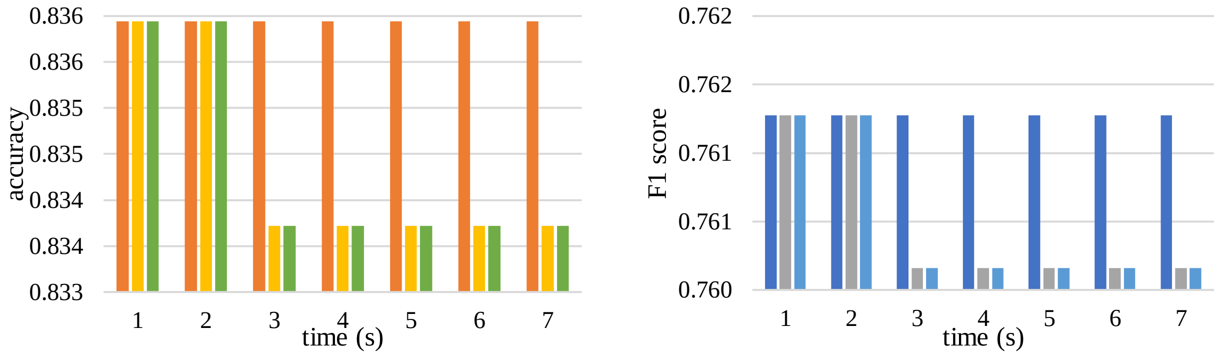
The cry detection model for babies that was constructed in this study utilizes the SVM algorithm. The selection of the Radial Basis Function (RBF) kernel for this model was based on its effectiveness in audio classification and multiclass classification scenarios [10], [15]. The chosen parameters for



(a) Accuracy

(b) F1 score

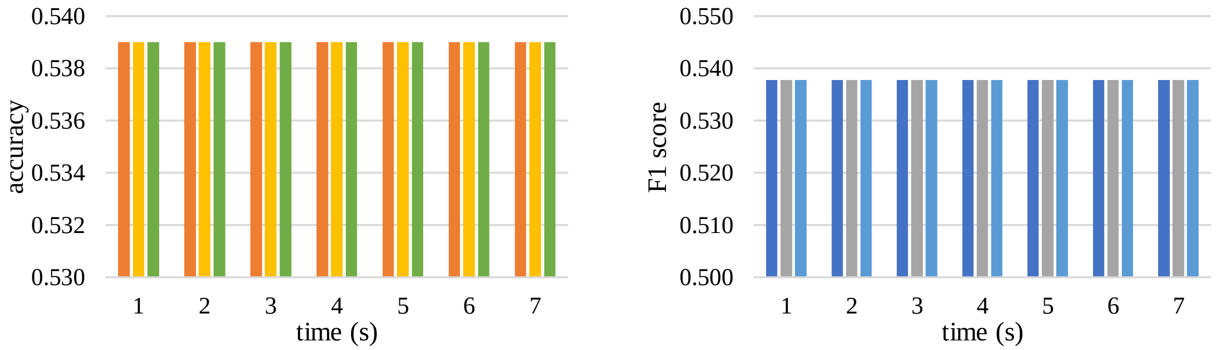
FIGURE 4 Random Forest performance using 10, 20, 30 MFCC



(a) Accuracy

(b) F1 score

FIGURE 5 Support Vector Machine performance using 10,20 and 30 MFCC



(a) Accuracy

(b) F1 score

FIGURE 6 Naïve Bayes performance using 10,20 and 30 MFCC

the SVM model include C=1.0, gamma = 'scale', probability = True, and decision_function_shape = 'ovr'.

The performance of the SVM model can be observed in FIGURE 5, whereby part (a) illustrates the accuracy and part (b) displays the F1 score value. Upon comparing the performance in FIGURE 5, it becomes evident that the modifications made to the number of MFCC utilized in the SVM model yield nearly identical results. Specifically, the accuracy ranges between 0.834 and 0.836, while the F1 scores vary between 0.760 and 0.761. Remarkably, the model achieves optimal performance when utilizing only one second of audio data.

C. PERFORMANCE OF NAÏVE BAYES

The baby cry detection model built using the Naïve Bayes algorithm in this study uses parameters n_samples = 500, n_features = 10, n_informative = 5, n_classes = 5, random_state = 1. The model's detection performance is illustrated in FIGURE 6.

FIGURE 6 provides a performance comparison, revealing no discrepancy in accuracy values and F1 scores when considering different MFCC coefficient values and audio length. The model in question achieves an accuracy value of 0.539 and an F1 score of 0.538.

IV. DISCUSSION

From the research results above, the best performance value of the models built with three classification algorithms with variations in the MFCC coefficient value and audio length of the baby's cry is known. Such outcomes are presented in TABLE 8, which showcases the optimal performance of the models created by utilising SVM, Random Forest, and Naïve Bayes classification algorithms.

A comparative analysis of the three models' best performance is illustrated in FIGURE 7. This comparison reveals that the Random Forest algorithm model outperforms the remaining two models. Furthermore, FIGURE 7 demonstrates that the models constructed using SVM and Random Forest algorithms could be further developed for detecting emotions in baby crying audio. Conversely, the Naïve Bayes model does not demonstrate satisfactory performance in baby crying audio.

TABLE 8
Result in Different Machine Learning Method Classification

Machine Learning Method	F1	Accuracy
SVM	0.761	0.836
Random Forest	0.773	0.844
Naïve Bayes	0.538	0.539

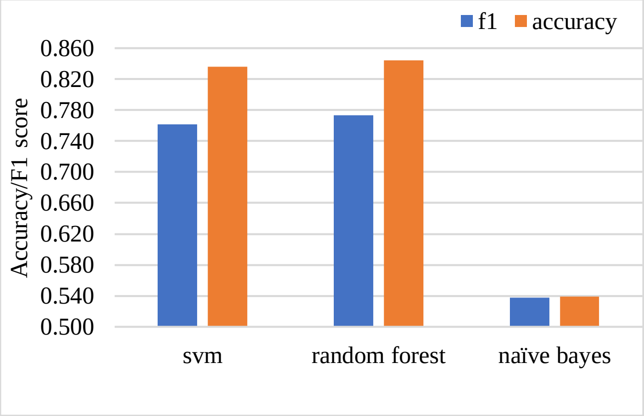


FIGURE 7. Comparison performance of machine learning methods.

The performance of the baby cry detection model built with Random Forest is influenced by the audio length and MFCC coefficient value. The influence of the two parameters is shown in FIGURE 4. The best performance of this model is obtained when using the smallest number of MFCC coefficients, namely ten and an audio length of six seconds.

Variations in the MFCC coefficient value do not affect the baby cry detection model built using the SVM and Naïve Bayes algorithms. FIGURE 5 and FIGURE 6 show the same model performance even though the MFCC coefficient value is changed. In these two figures, it can also be seen that the best model performance has been obtained using an audio length of one second.

TABLE 9 shows a comparison of the performance of the models built in this research with the performance of models from previous research. Previous studies used 7 seconds of audio of babies crying.

Research on baby crying has been carried out and uses the donate-a-cry-corpus dataset [23]. This research obtained accuracy results of 81.27%. These results were obtained with the Convolutional Neural Network (CNN) model. Other research also uses datasets and MFCC-based feature extraction, and several classification algorithms [12]. The algorithms used in this research are Random Forest, K Nearest Neighbors (KNN), SVM and Linear Regression (LR). The audio length of the baby's cry used in this study was 7 seconds. The best performance of the research used a model built with Random Forest with an accuracy value of 84%.

TABLE 9
Previous Research On Baby cry Dataset

Research	Classifier	Accuracy (%)
[23]	CNN	81.2
[12]	**Random Forest**	**84**
	KNN	82
	SVM	71
	LR	42
Our Research	**Random Forest**	**84.444**
	SVM	83.590
	Naïve Bayes	53.900

A comparison between the two preceding studies reveals that the method proposed in this study has the potential to outperform or achieve similar results as previous research. The method suggested in this research exhibits a distinct advantage: it yields superior outcomes when utilizing shorter audio durations. The investigation also introduces novel aspects, such as the examination of variations in MFCC coefficient values and audio length, which have not been extensively explored in the realm of audio classification. This is particularly true for research of the classification of baby's cries.

Nevertheless, it is important to acknowledge the limitations and deficiencies within this research. Specifically, the resulting model's performance fails to reach optimal levels, as indicated by an accuracy rate below 85% and an F1 score below 0.8. The suboptimal performance of the baby cry detection can potentially be attributed to the issue of unbalanced data. Additionally, the audio segmentation process solely considers duration, disregarding the actual content. Consequently, the extracted audio snippets may either lack sound or contain non-cry sounds. If such audio data is utilized during the training phase, it is plausible that the resultant model will exhibit suboptimal performance.

The cry detection model for babies developed in this study will have significant implications in the healthcare industry due to its potential utilization by medical professionals and parents. Implementing this detection model as a mobile application would enable medical professionals to decipher the underlying issues causing the baby's needs based on their cries. Furthermore, parents would also benefit from this innovation as it would allow them to easily comprehend their babies' emotions and needs. Additionally, this research has profound implications in computer science as it contributes to the advancement of knowledge in audio classification research in general, specifically in the classification of infant cries.

V. CONCLUSION

The data obtained from baby crying audio is considered unstructured and requires a feature extraction procedure to generate structured data suitable for machine learning algorithms. In this study, Mel Frequency Cepstral Coefficients (MFCC) served as the basis for the feature extraction technique, with varying coefficient values employed to process baby crying audio spanning 1 to 7 seconds. This investigation encompassed the development of three detection models utilizing three distinct machine learning algorithms: Support Vector Machine (SVM), Random Forest, and Naive Bayes, resulting in respective accuracy rates of 0.836, 0.844, and 0.539. Additionally, the F1 score values for the models above were calculated as 0.761, 0.773, and 0.538, respectively.

This research still has several limitations if seen from the model's performance, which produces an F1 score below 0.8. This suboptimal model performance can be caused by the model being built using unbalanced data. As a result, predictions for minority class data are wrong because the model tends to follow the pattern of majority class data. Another limitation is that audio cutting is only done based on the desired duration without paying attention to the content. This method has the potential to produce audio files whose contents are empty, making the pattern recognition training process less accurate.

Given these limitations and shortcomings, it is recommended that further research be conducted to acquire new data that includes minority class data, thus achieving a balanced dataset. Additionally, future investigations should explore the implementation of data balancing techniques to enhance baby cry detection performance.

VI. ACKNOWLEDGMENT

We sincerely thank all the individuals from the Computer Science Department, Lambung Mangkurat University, and the Information System Department, Nahdlatul Ulama Surabaya University, who have contributed to the successful completion of this computer project. Their valuable inputs and suggestions have significantly improved the quality of our work. We are also indebted to our project team members for their collaboration and hard work, which have been vital to the research's accomplishments.

REFERENCES

[1] C. Ji, T. B. Mudiyanselage, Y. Gao, and Y. Pan, "A review of infant cry analysis and classification," *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2021, no. 1, 2021, doi: 10.1186/s13636-021-00197-5.

[2] S. Mishra, "Artificial Intelligence: A Review of Progress and Prospects in Medicine and Healthcare," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 4, no. 1, pp. 1–23, 2022, doi: 10.35882/jeeemi.v4i1.1.

[3] D. F. Sengkey and A. S. R. Masengi, "Regression Algorithms in Predicting the SARS-CoV-2 Replicase Polyprotein 1ab Inhibitor: A Comparative Study," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 1, pp. 1–10, 2024, doi: 10.35882/JEEEMI.V6I1.338.

[4] K. Sharma, C. Gupta, and S. Gupta, "Infant Weeping Calls Decoder using Statistical Feature Extraction and Gaussian Mixture Models," *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, pp. 1–6, 2019, doi: 10.1109/ICCCNT45670.2019.8944527.

[5] F. Anders, M. Hlawitschka, and M. Fuchs, "Automatic classification of infant vocalization sequences with convolutional neural networks," *Speech Communication*, vol. 119, no. October 2019, pp. 36–45, 2020, doi: 10.1016/j.specom.2020.03.003.

[6] P. Sandhya, V. Spoorthy, S. G. Koolagudi, and N. V. Sobhana, "Spectral Features for Emotional Speaker Recognition," *Proceedings of 2020 3rd International Conference on Advances in Electronics, Computers and Communications, ICAECC 2020*, 2020, doi: 10.1109/ICAECC50550.2020.9339502.

[7] D. Berrar, "Cross-validation," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, no. January 2018, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.

[8] L. Le, A. N. M. H. Kabir, C. Ji, S. Basodi, and Y. Pan, "Using Transfer Learning, SVM, and Ensemble Classification to Classify Baby Cries Based on Their Spectrogram Images," *Proceedings - 2019 IEEE 16th International Conference on Mobile Ad Hoc and Smart Systems Workshops, MASSW 2019*, pp. 106–110, 2019, doi: 10.1109/MASSW.2019.00028.

[9] F. Salehian Matikolaie and C. Tadj, "On the use of long-term features in a newborn cry diagnostic system," *Biomedical Signal Processing and Control*, vol. 59, p. 101889, 2020, doi: 10.1016/j.bspc.2020.101889.

[10] M. M. Mafazy, "Classification of COVID-19 Cough Sounds using Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction and Support Vector Machine Telematika Classification of COVID-19 Cough Sounds using Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction," no. August, 2023, doi: 10.35671/telematika.v16i2.2569.

[11] S. Mishra, "A Comparative Study for Time-to-Event Analysis and Survival Prediction for Heart Failure Condition using Machine Learning Techniques," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 4, no. 3, pp. 115–134, 2022, doi: 10.35882/jeeemi.v4i3.225.

[12] P. Kulkarni, S. Umarani, V. Diwan, V. Korde, and P. P. Rege, "Child Cry Classification - An Analysis of Features and Models," *2021 6th International Conference for Convergence in Technology, I2CT 2021*, pp. 1–7, 2021, doi: 10.1109/I2CT51068.2021.9418129.

[13] A. Ekİncİ and E. Küçükkülahli, "Classification of Baby Cries Using Machine Learning Algorithms," vol. IX, no. I, pp. 16–26, 2023.

[14] I. Södergren, M. P. Nodeh, P. C. Chhipa, K. Nikolaidou, and G. Kovács, "Detecting COVID-19 from audio recording of coughs using Random Forests and Support Vector Machines," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 6, no. November, pp. 4256–4260, 2021, doi: 10.21437/Interspeech.2021-2191.

[15] K. Rezaee, H. Ghayoumi Zadeh, L. Qi, H. Rabiee, and M. R. Khosravi, "Can you Understand why I am Crying? A Decision-making System for Classifying Infants' Cry Languages Based on deepSVM Model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2023, doi: 10.1145/3579032.

[16] R. I. Tuduce, M. S. Rusu, H. Cucu, and C. Burileanu, "Automated baby cry classification on a hospital-acquired baby cry database," *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*, pp. 343–346, 2019, doi: 10.1109/TSP.2019.8769075.

[17] G. Aggarwal, K. Jhajharia, J. Izhar, M. Kumar, and L. Abualigah, "A

Machine Learning Approach to Classify Biomedical Acoustic Features for Baby Cries," *Journal of Voice*, Jul. 2023, doi: 10.1016/J.JVOICE.2023.06.014.

[18] M. R. Faisal *et al.*, "LSTM and Bi-LSTM Models For Identifying Natural Disasters Reports From Social Media," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 4, 2023.

[19] S. Aini, W. A. Kusuma, M. K. D. Hardhienata, and Mushthofa, "Network-Based Molecular Features Selection to Predict the Drug Synergy in Cancer Cells," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 3, pp. 168–176, 2023.

[20] V. Bansal, G. Pahwa, and N. Kannan, "Cough classification for COVID-19 based on audio mfcc features using convolutional neural networks," *2020 IEEE International Conference on Computing, Power and Communication Technologies, GUCON 2020*, pp. 604–608, 2020, doi: 10.1109/GUCON48875.2020.9231094.

[21] A. C. Kemila, W. Fawwaz, and A. Maki, "Parameter Optimization of Support Vector Machine using River Formation Dynamic on Brain Tumor Classification," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 3, pp. 177–184, 2023.

[22] M. T. Hidayat, M. R. Faisal, D. Kartini, F. Indriani, and I. Budiman, "Comparison of Machine Learning Performance on Classification of COVID-19 Cough Sounds Using MFCC Features Porównanie wydajności uczenia maszynowego w zakresie klasyfikacji odgłosów kaszlu COVID- 19 przy użyciu funkcji MFCC," vol. 29, no. August, pp. 399–404, 2023.

[23] E. Sutanto, F. Fahmi, W. Shalannanda, and A. Aridarma, "Cry Recognition for Infant Incubator Monitoring System Based on Internet of Things using Machine Learning," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 444–454, 2021, doi: 10.22266/IJIES2021.0228.41.

[24] R. T. Yunardi, R. Apsari, and M. Yasin, "Comparison of Machine Learning Algorithm For Urine Glucose Level Classification Using Side-Polished Fiber Sensor," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 2, no. 2, pp. 33–39, 2020, doi: 10.35882/jeeemi.v2i2.1.

[25] M. Anbu and G. S. Anandha Mala, "Feature selection using firefly algorithm in software defect prediction," *Cluster Computing*, vol. 22, no. 4, pp. 10925–10934, 2019, doi: 10.1007/s10586-017-1235-3.

[26] J. He, L. Yang, D. Liu, and Z. Song, "Automatic Recognition of High-Density Epileptic EEG Using Support Vector Machine and Gradient-Boosting Decision Tree," *Brain Sciences*, vol. 12, no. 9, 2022, doi: 10.3390/brainsci12091197.

[27] M. I. Mazdadi, I. Budiman, and R. Herteno, "Implementation of Information Gain and Particle Swarm Optimization on Sentiment Analysis of Covid-19 Handling Using K-Nn," *Jurnal Informatika dan Komputer) Accredited KEMENDIKBUD RISTEK*, vol. 6, no. 1, pp. 261–270, 2023.

[28] M. R. Ansyari, M. I. Mazdadi, D. Kartini, and T. H. Saragih, "Implementation of Random Forest and Extreme Gradient Boosting in the Classification of Heart Disease Using Particle Swarm Optimization Feature Selection," vol. 5, no. 4, 2023.

[29] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[30] T. T. A. Putri, S. Sriadhi, R. D. Sari, R. Rahmadani, and H. D. Hutahaean, "A comparison of classification algorithms for hate speech detection," *IOP Conference Series: Materials Science and Engineering*, vol. 830, no. 3, 2020, doi: 10.1088/1757-899X/830/3/032006.

BIBLIOGRAPHY

Putri Agustina Riadi originated in Banjarbaru, South Kalimantan. Since 2020, she has pursued her academic endeavors as a student of the Computer Science Department at Universitas Lambung Mangkurat. Her current area of research lies within the realm of data science. The study program offers the opportunity to cultivate her interest in data science. She has selected this particular interest due to my affinity towards data science and her profound fascination with this field. Additionally, her final project entailed conducting research that centered around the classification of a baby's audio machine-learning method. The purpose of this research endeavor is to determine the influence of the audio period of a baby cry on machine learning classification results.

Mohammad Reza Faisal was born in Banjarmasin. Following his graduation from high school, he pursued his undergraduate studies in the Informatics department at Pasundan University in 1995 and later majored in Physics at Bandung Institute of Technology in 1997. After completing his bachelor's program, he gained experience as a training trainer in the field of information technology and software development. Since 2008, he has been a lecturer in computer science at Universitas Lambung Mangkurat, while also pursuing his master's program in Informatics at Bandung Institute of Technology in 2010. In 2015, he furthered his education by pursuing a doctoral degree in Bioinformatics at Kanazawa University, Japan. To this day, he continues his work as a lecturer in Computer Science at Universitas Lambung Mangkurat. His research interests encompass Data Science, Software Engineering, and Bioinformatics.

Dwi Kartini received her Bachelor's and master's degrees in computer science from the Faculty of Computer Science, Putra Indonesia "YPTK" Padang, Indonesia. She is a lecturer too in the Department of Computer Science. She instructing in various subjects such as linear algebra, discrete mathematics, research methods and others Her research interests include the applications of Artificial Intelligence and Data Mining. She is an assistant professor in the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University in Banjarbaru, Indonesia.

Radityo Adi Nugroho received his bachelor's degree in Informatics from the Islamic University of Indonesia and a master's degree in Computer Science from Gadjah Mada University. Currently, he is an assistant professor in the Department of Computer Science at Lambung Mangkurat University. His research interests include software defect prediction and computer vision. He is also a practitioner in the field of information technology as a project manager and systems analyst to develop software and information systems used by universities

Dodon Turianto Nugrahadi is a lecturer in Department of Computer Science, Lambung Mangkurat University. His research interest is centered on Data Science and Computer Networking. He completed his bachelor's degree in Informatics Engineering in the UK. Petra, Surabaya in 2004. After that, he pursued a master's degree in Information Engineering at Gajah Mada University, Yogyakarta in 2009. His current area of research revolves around Network, Data Science, Internet of Things (IoT), and network Quality of service (QoS).

Dike Bayu Magfira holds a Bachelor's degree in Informatics Engineering from Bengkulu University and a Master's degree in Technology Management from the Sepuluh November Institute of Technology, Surabaya. Her research interests include artificial intelligence and software development. From 2019 to early 2022, she taught at several campuses in the city of Surabaya as a practitioner lecturer for software design and software project management courses because she was still working at a software development company. Since 2022, she has been a permanent lecturer at Nahdlatul Ulama University, Surabaya, in the information systems study program.