Audio Feature Analysis and Selection for Deception Detection in Court Proceedings

Muhammad Meftah Mafazy [1)], Chastine Fatichah [2,*)], and Anny Yuniarti [3)]

[1, 2, 3)] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: meftah.mafazy@gmail.com[1)], chastine@its.ac.id[2)], and anny@its.ac.id[3)]

ABSTRACT

Deception detection is a method to determine whether a person is lying or not. One lie detector is a polygraph that measures human physiology, such as pulse and blood pressure. However, polygraphs have a problem in that they cannot be measured based on human psychology, such as speech and intonation. Therefore, audio deception detection is required, and this can be measured based on human psychology. This research will extract audio features, such as the Mel Frequency Cepstral Coefficient (MFCC), Jitter, Fundamental Frequency (F0), and Perceptual Linear Prediction (PLP), from the Real-Life Trial dataset, which comprises 121 audio data. From the extraction results in the form of numerical data totaling 6387 features, various feature-selection methods are employed, such as Feature Importance (FI), Principal Component Analysis (PCA), Information Gain, Chi-Square, and Recursive Feature Elimination (RFE). After feature selection, the selected features are input to machine learning models, such as random forest and support vector machine (SVM). After model testing, metrics such as accuracy, precision, recall, and F1 score were evaluated, as well as statistical evaluation, to assess the developed model. Results from this experiment show that the deception detection model is improved after a feature selection process to reduce irrelevant features. Comparing the accuracy, Chi-Square achieves a significantly higher result, reaching up to 92% with an improvement of 24.32%, surpassing the SVM model's accuracy of 67.57% before feature selection. In contrast, the RFE technique yielded the best accuracy of 86%, with an increase of 13.52%, building upon its baseline accuracy of 72.97%.

Keywords: Deception detection, feature extraction, feature selection, random forest

1. Introduction

Lying conveys false information intended to deceive others, usually by stating something that is not true [1]. Lying can be characterized by multiple definitions, including supplying inconsequential data, altering data, and misrepresenting information [2]. Deception can be detected using reliable methods to determine whether someone is lying. Deception detection methods are primarily used in criminal inquiries, which seek to establish the truth of original statements or false testimonies provided by witnesses or defendants during trial proceedings [3]. The device commonly used to identify deception is a physiological polygraph. The polygraph is usually used to measure physiological responses, including blood pressure, pulse rate, and breathing pattern [4]. Despite their potential, polygraph have a limitation in that they cannot accurately distinguish between deception and other psychological nonverbal cues, including speech patterns, lip movements, intonation, and body language [5]. From this perspective, scientists have conducted extensive research on deception detection grounded in human psychology, one of which uses audio media to extract features, changing, the form of audio data into numerical data that represent the audio.

Feature extraction is an important aspect of audio-based research. The purpose of this extraction is to process previously acquired audio data and the extract numerical features representing the audio data itself [6]. There are many audio feature extraction function and methods available, such as Mel-Frequency Cepstral Coefficient (MFCC), which represent the cepstral coefficient of each audio data with the aim of distinguishing features of characteristics [7]. In addition to MFCC, there is a Jitter feature that generates numerical data based on changes in sound vibrations. This feature can be used to identify a person's stress or anxiety level [8]. Then there is also

---

the extraction of Fundamental Frequency (F0) features that represent the analysis of sound signals to recognize individual characteristics in the detection of deception, while other studies also mention that people who are angry or feel anxious tend to have a higher fundamental frequency with greater fluctuations. In addition, people who experience depression typically exhibit a lower fundamental frequency of changes [9]. In addition to MFCC, Jitter, and F0 extraction, there is Perceptual Linear Prediction (PLP) feature extraction, which is a derivative of Linear Prediction Coefficients (LPC) by using the psychophysical concept of human hearing to estimate the auditory spectrum [10]. The use of feature extraction techinques, such as MFCC, Jitter, F0, and PLP, provides various important information that helps recognize relevant patterns. However, for a deception detection model to work optimally, a strategy is required to ensure that only relevant features are used.

The main challenge in training machine learning models for deception detection is the complexity of datasets such as high-dimensional data [11]. High-dimensional data often contain redundant and even irrelevant features that can affect the training results of machine learning models, leading to overfitting, increased computational time, and reduced prediction accuracy of machine learning models [12]. To overcome these problems, feature selection is required to remove unnecessary features. Feature selection itself has been proven to improve the accuracy and speed up of machine learning model training [13]. Following feature selection, the subsequent step is to construct a deception detection model. Research has shown that machine learning models like Random Forest and SVM are effective in identifying deception, as evidenced by studies such as Li et al.'s work on detecting deception in videos through the use of wrapper-based feature selection and machine learning techniques. The study achieved its highest accuracy of 83.33% with the SVM method, exceeding the 78.50% accuracy of the KNN method [5]. Another study by Sen et al. also investigated the detection of multimodal-based deception, including a study focusing on audio-based methods. The study also compares various machine learning models, including one that employs Random Forest. According to the study's findings, a Random Forest achieved the highest level of accuracy at 63.28%, surpassing other methods such as Neural Network, which had an accuracy rate of 61.02% [14].

Machine learning is widely used in various cases, especially in deception detection cases. In addition to many uses, the variety of machine learning methods has also varied, one of which is the method created by Leo Breiman in 2001, Random Forest [15]. In addition, Random Forest is an ensemble-based algorithm that is quite reliable for handling datasets with high dimensions and overfitting, and it can produce less accurate accuracy than other classification techniques [16]. In addition to Random Forest, there is another method, namely, Support Vector Machine (SVM), which is known as a machine learning method that represents data as points and projects them into an n-dimensional space where the data are classified into groups separated by a hyperplane. The hyperplane was constructed by maximizing the distance or margin between the feature space and the nearest data point [17]. In another study, Nasri et al. developed an audio-based deception detection model that employs an SVM. The data used is interview data collected from several volunteers such as students. Each student was asked to answer prepared and unexpected research questions. The study also extracted features such as MFCC and Pitch first because the data were still in the audio format. Based on the MFCC and Pitch extraction results, model testing was performed using SVM. The accuracy of the results was approximately 90% [18].

In this study, the audio features MFCC, Jitter, F0, and PLP were extracted. Because it indicates the identity of the voice signal, MFCC is essential in audio studies. In addition, Jitter is important because it conveys information about the intensity and vibration of the vocal cords and distinguishes between deceptive and truthful speech. Furthermore, F0 correlates with anger and fear emotions associated with lying. Thus, the PLP takes into account the quality of human hearing and extracts relevant information from the voice input, so it works well in speech recognition and deception detection. Voice quality and spectral features are used in feature extraction. However, strong correlation and redundancy can slow recognition and extend calculation time. To reduce useless features and minimize calculation time, feature selection methods such as Feature Importance, Principal Component Analysis (PCA), Information Gain, Chi-Square, and Recursive Feature Elimination (RFE), are explored. Then, after exploring feature selection, machine learning models, such as Random Forest and SVM models, will be tested. The

experimental results are clearly visible from the feature selection exploration and machine learning models. For model evaluation, we used metric evaluations, such as accuracy, precision, recall, and f1-score.

2. Related Works

Research related to detecting deception has started to gain significant attention from numerous researchers, particularly within the context of court trials. At a trial, deception detection can serve as a valuable aid in determining whether a witness or defendant is being dishonest. Despite their physiological basis, polygraphs have limitations, primarily their inability to gauge human psychological responses, including behaviour, vocal tone, and mouth movements. Previous research has employed different feature extraction methods and machine learning models for deception identification, as seen in a study by Srivastava et al., who successfully identified deceitful behaviour by utilising a machine learning methodology. The initial features extracted from the subject's voice in the study were Mel Frequency Cepstral Coefficient (MFCC), Zero Crossing Rate (ZCR), Energy, and Fundamental Frequency (F0). Alongside the voice recordings, the researchers conducted psychological assessments of the participants, comprising heart rate, blood pressure, and respiratory measures. The results show that the SVM algorithm achieved the highest accuracy of 100%, outperforming the Artificial Neural Network (ANN) which secured an accuracy of 93% [19].

In addition, other research conducted by Javaid et al. detected lies through various aspects, one of which is audio using MFCC feature extraction, Linear Prediction Coefficient (LPC), Perceptual Linear Prediction (PLP) and Discrete Wavelet Transform (DWT). The study used a bag-of-lies dataset consisting of several aspects, one of which was audio data with 325 recorded data consisting of 35 subjects. From the experiment, an accuracy of 80.6% was obtained using a Convolutional Neural Network (CNN) for the audio aspect [20]. Similar research conducted by Wang et al. detected lies through machine learning-based audio using data collected from several students at a university in China. From the data collection, 890 audio data were obtained using a division of 439 and 451 data, which were labeled as lies and 451 data labeled as honesty. Then, feature extraction is performed using MFCC, Jitter, and F0 along with several machine learning algorithms, one of which is Random Forest. The results demonstrate an accuracy of 91.52% when using Random Forest [9]. Some research results have demonstrated that audio features such as MFCC, Jitter, F0, and PLP, have potential in the deception detection domain. However, the selection of relevant features is a major challenge to improve accuracy and prevent overfitting; thus, feature selection explorations such as Feature Importance, Principal Component Analysis (PCA), Information Gain, Chi-Square, and Recursive Feature Elimination (RFE), are required to optimize classification using machine learning such as Random Forest and Support Vector Machine (SVM).

The use of feature selection by researchers has been widely used in classification and regression cases, especially in deception detection, to improve machine learning models. In Gharsalli et al.'s research on human expression detection using Feature Importance feature selection, Gharsalli et al. obtained 92.160 features from 123 facial expression video data. Then, features were selected using Feature Importance, resulting in 500 selected features. Classification was performed using a Support Vector Machine (SVM), and a noticeable improvement was observed for 90% of the selected values, which previously only reached 83%. The results of this study highlight the capability of feature selection techniques in machine learning algorithms [21]. Further research conducted by Al-Dhaher et al. [22] revealed additional evidence on deception detection. The focus of this investigation was on the different characteristics of noisy audio, including the mel frequency cepstral coefficient, jitter, and Base Frequency (F0). The results showed an improvement of 88% using feature importance feature selection combined with a random forest classification method that included 30 features [22]. Principal Component Analysis (PCA) can also be used in audio-based deception detection, as demonstrated by Fernandes et al. using features such as Time Difference Energy, Delta Energy, Time Difference Cepstrum, and Delta Cepstrum, and these features are evaluated with Levenberg-Marquardt and Long Short Term Memory (LSTM) algorithms. The number of audio features was reduced by PCA. The results of this study were almost 100% accurate, indicating that PCA can improve the reliability of deception detection models [23]. In addition to Feature Importance and PCA, feature selection by Information Gain can also be used, as demonstrated by Liu et al., who investigated linguistic features for deception

detection. Prior to deception detection, Information Gain feature selection was performed to select relevant features after the text data extraction process. The proposed method achieved an accuracy rate of 81% when using the SVM [24].

Other studies have demonstrated that the chi-square approach can improve the accuracy of machine learning models, as demonstrated by Nishi et al., who used machine learning to identify credit card fraud. This investigation began with a feature selection process involving chi-square, which was used to assess the efficacy of the 4 machine learning models. The results of this study demonstrate that logistic regression is the most effective approach for detecting credit card fraud, with a success rate of 98.90% [25]. Recursive Feature Elimination (RFE) is a feature selection technique used in classification or regression cases, such as deception detection. The proposed method can also be compared to other methods, such as feature importance, PCA, information gain, and chi-square. Shukla et al. used RFE and a dynamic ensemble approach to detect credit card fraud and achieved 90% accuracy [26]. The feature selection that produces the selected features is continued into the design stage of the deception detection model using machine learning.

In general, machine learning techniques, such as random forest and SVM, have been widely implemented in various cases of classification or detection, such as the research conducted by Kumara et al. to develop audio-based deception detection to assess customer eligibility in paying loans. This research first extracts audio features such as MFCC, Chroma, Spectral Centroid, Spectral Bandwidth, Spectral Contrast, Chroma Short-Time Fourier Transform (CSTFT), and ZCR. This research uses machine learning algorithms such as Random Forest. From this research, obtained an accuracy of 71.45% when using Random Forest [27]. Another study by Yang et al. investigated deception detection based on emotion level recognition. This study used a public dataset of 121 videos. From the extraction results, classification was performed using several machine learning algorithms, one of which was SVM, which was obtained with an accuracy of 87.59% [28].

Research related to deception detection has shown significant development annually using various approaches, such as the use of feature extraction and machine learning algorithms for classification. Various feature extraction approaches such as MFCC, Jitter, F0, and PLP, have also been considered to improve deception detection accuracy. In addition, the use of feature selection such as feature importance, PCA, information gain, chi-square, and RFE, is also considered to optimize machine learning models. In addition to feature extraction and selection, the use of machine learning models, such as random forest and SVM, can improve deception detection models.

3. Methodology

Here, we describe the research flow plan shown in Fig. 1, which represents the stages of the deception detection research. The first step is to explore and collect deception detection datasets. Then, after collecting the data, we proceed with the preprocessing stage of the dataset, such as changing the shape of the data. Subsequently, previously described feature extraction experiments were carried out such as MFCC, Jitter, F0, and PLP. From the extraction results, which will be in the form of numerical data, various feature selection methods are performed, such as Feature Importance, Principal Component Analysis (PCA), Information Gain, Chi-Square, and Recursive Feature Elimination (RFE) to identify which features are relevant enough. After feature selection, the selected features are used as input for model design and testing using random forest and support vector machines, which also optimize the hyperparameters of each machine learning method. After model testing, metrics such as accuracy, precision, recall, and the f1 score are evaluated as well as statistical evaluation to assess whether the developed model is good or there is still something that needs to be improved.

3.1. Collecting Data

The Real-Life Trial dataset from the University of Michigan-Deaborn was used in this study. This dataset contains a collection of videos during a real-life trial containing statements made by ex-convicts after being released, and there are several statements from the defendant on one of the television shows related to crime. In addition, the speaker of this video dataset is the defendant or witness. Then, these video data amount to 121 video data with 2 labels, namely, deceptive and truthful [29]. Fig. 2 presents an overview of the Real-Life Trial dataset.
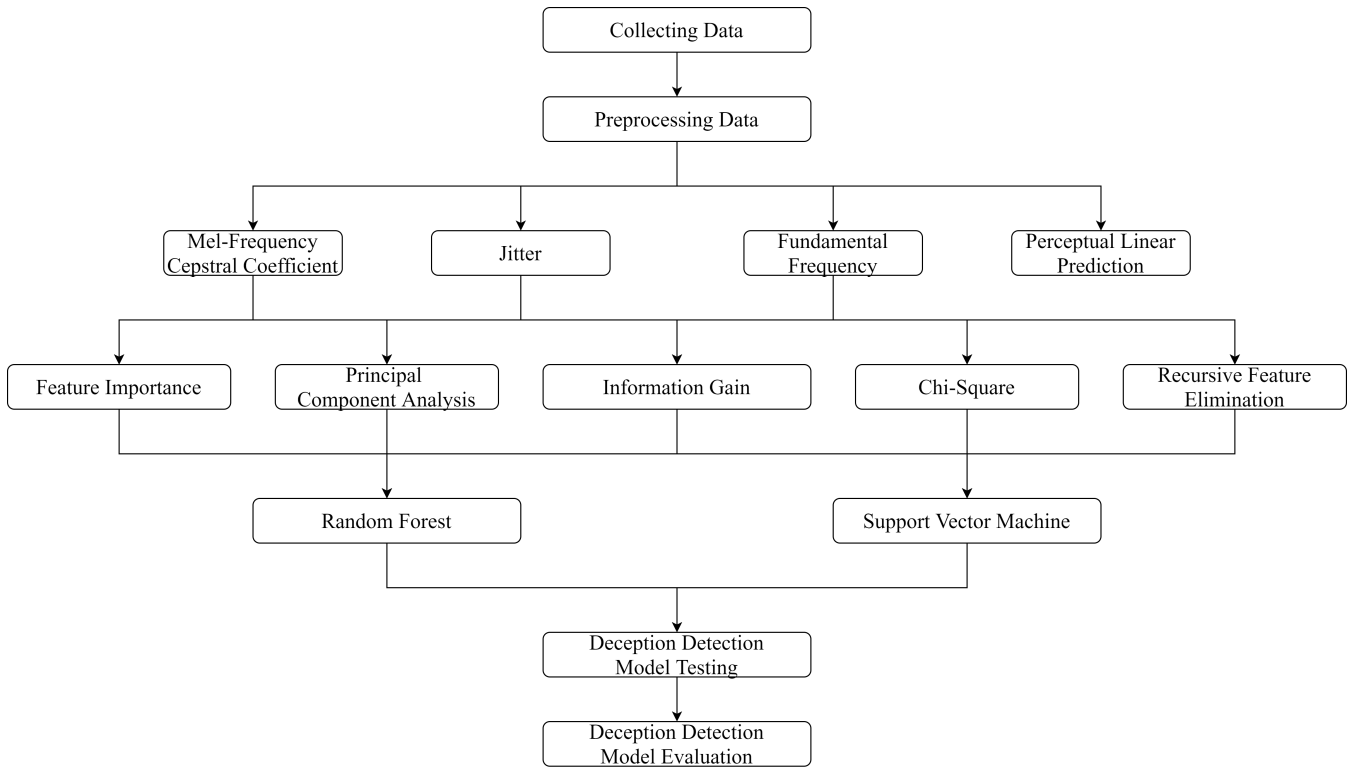
Fig. 1: Research Flow

Fig. 2: Real-Life Trial Dataset [30]

3.2. Preprocessing Data

In this research, the data preprocessing stage is needed first, because the Real-Life Trial dataset is a video dataset. The first process is to change the format of the dataset which was previously video in the form of .mp4 will be converted into audio data in the form of .wav. The process of changing the format of this dataset simply uses a library from the Python programming language, Moviepy. Fig. 3 is an illustration of video data that has been converted into audio data in the form of a spectrogram representation.

Each dataset labelled as deceptive and truthful in Fig. 3 is analyzed via a spectrogram. The sound analysis depicted in the spectrogram pattern indicates deception. Generally, the pattern exhibited by an individual who is being dishonest typically displays a flat signal pattern, with the predominance of black color being a clear indicator, suggesting several potential explanations including heightened noise levels, low vocal intensity, and emotional instability, such as anxiety, when conveying information. The spectrogram in Fig. 4 corresponds to a sample from the Real-Life Trial dataset that has been labeled as deceptive.

Furthermore, the spectrogram pattern of individuals speaking truthfully is closely similar to that of the spectrogram identified as indicative of deception. Viewed from the perspective of the signal color, a straightforward pattern exhibits a more prominent light hue, suggesting the potential for the speaker to communicate information
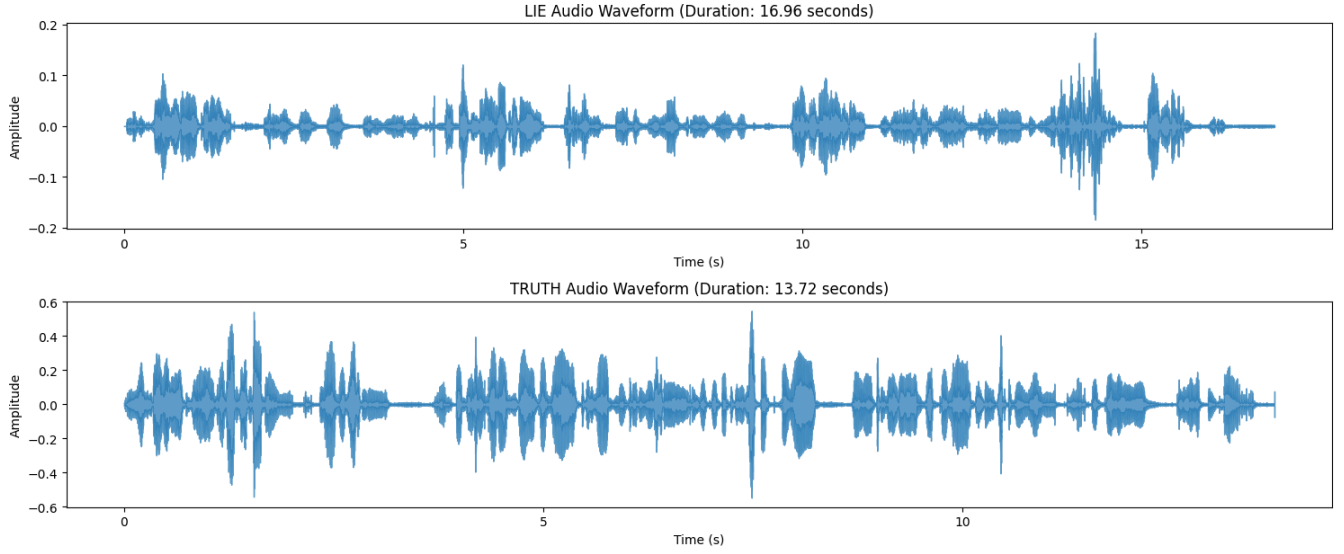
Fig. 3: Real-Life Trial dataset samples that have been converted into audio data
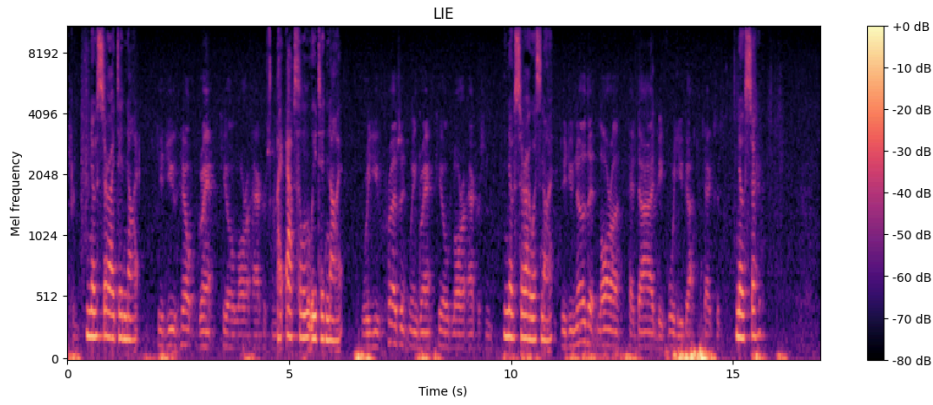

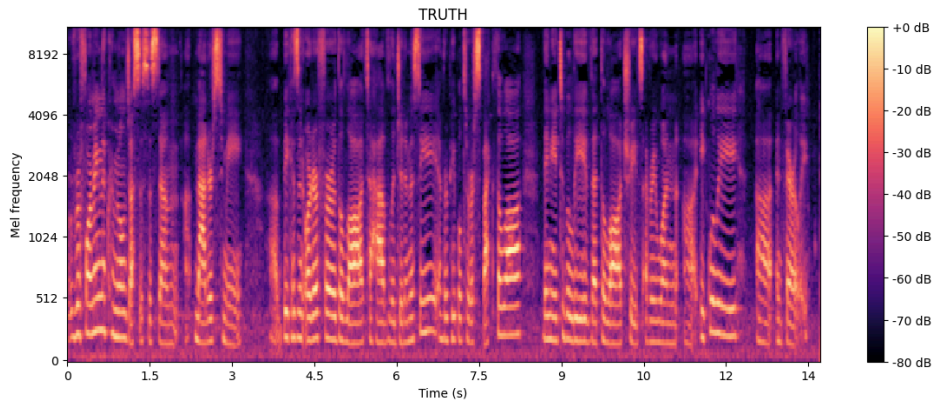Fig. 4: Spectrogram on Real-Life Trial dataset samples labeled as deceptive

Fig. 5: Spectrogram on Real-Life Trial dataset samples labeled as truthful

clearly, accurately, and in a direct manner. The spectrogram in Fig. 5 corresponds to the truthful sample from the Real-Life Trial dataset.

3.3. Feature Extraction

MFCC, Jitter, F0, and PLP feature extraction are numerical data extracted from audio data previously in the form of videos representing each audio data. This research will use the openSMILE and Shennong feature extraction libraries, which are already available in the Python programming language. OpenSMILE will extract MFCC, Jitter, and F0 features. Meanwhile, Shennong will extract PLP features.

A. Mel Frequency Cepstral Coefficient (MFCC)

MFCC has been widely used, especially in deception detection. The MFCC extraction process begins with pre-emphasis to equalize the signal spectrum of the voice itself. Then, the pre-emphasis signal that has been processed by the spectrum, will break down into several overlapping frames which are then processed into the hamming windowing section to reduce the edge effect. After that, the frames that have gone through the hamming windowing process will be transformed again to frequency using the Fast Fourier Transform (FFT) [6]. The results of this transformation will later be converted into a mel-frequency scale to simulate the human auditory response. In this research, the MFCC will be extracted as many as 14 coefficients in the openSMILE library.

B. Jitter

Jitter is one of the feature extractions that measures the instability of vocal fold vibrations contained in audio data. The instability is often referred to as micro-tremor. In addition, the jitter feature is quite crucial in the case of deception detection, because high jitter is related to the level of stress or anxiety when someone is lying [22]. In addition, jitter can also be used as a feature extraction that can assess overall voice quality [31]. In this research, jitter will be extracted as 2 coefficients in the openSMILE library.

C. Fundamental Frequency (F0)

F0 is known as the lowest frequency in a sound. F0 is usually used in human speech recognition cases such as detecting speakers or detecting diseases of the sound-producing organs because F0 can analyze sound signals. In addition to recognizing individual characteristics, fundamental frequency is also often used in music cases such as music detection and audio search [10]. In this research, the fundamental frequency will be extracted as much as 1 coefficient in the openSMILE library.

D. Perceptual Linear Prediction (PLP)

PLP is one of the feature extractions known as sound analysis designed to match human auditory perception. In other words, PLP is better at capturing the characteristics of the sounds heard. In addition, PLP also has the advantage of capturing emotion-relevant sound characteristics, such as intonation and frequency spectrum which will be useful for speech recognition cases such as deception detection [10]. In this research, PLP will be extracted as many as 13 coefficients in the Shennong library.

3.4. Feature Selection

Feature extraction processes such as MFCC, Jitter, F0, and PLP attempt to convert audio data into numerical data that represents the signals in the audio. Usually from the feature extraction process, there are features that may later be irrelevant to the research to be carried out, for this reason it is necessary to do feature selection. Basically, feature selection aims to find out the features that are relevant and influential in the case of deception detection itself. With feature selection, it is expected to improve model performance and reduce complexity in deception detection. In this research, we will focus on several feature selection methods such as Feature Importance, Principal Component Analysis (PCA), Information Gain, Chi-Square, and Recursive Feature Elimination (RFE).

A. Feature Importance

Random Forest is an ensemble algorithm that is quite often used for classification and regression, especially in deception detection. In addition to its classification capabilities, Random Forest can also perform feature selection by building many decision trees and measuring the importance of each feature using the Gini index with the aim of identifying the most contributing features. Then, the result of the feature measurement process is called feature importance[32]. After that, the features with the highest value resulting from the extraction of MFCC, Jitter, F0, and PLP will be used to build a machine learning model in audio-based deception detection later so that the computation run is not too large.

B. Principal Component Analysis (PCA)

PCA is a dimensionality reduction method technique that aims to identify the main direction of variation in the data [33]. The use of PCA in this research aims to overcome the problem of audio data that often has high

dimensionality from extracted audio features such as MFCC, Jitter, F0, and PLP which can later affect the efficiency of the classification process. By using PCA, it is expected to eliminate noise in audio data that has been extracted into numerical data and increase accuracy in deception detection.

C. Information Gain

Information Gain is one of the entropy-based feature extractions that is often used in various studies. Information Gain is usually used to recognize the most informative or crucial features in a dataset, improve accuracy and reduce the dimensionality of the dataset. This method evaluates the reduction of uncertainty after dividing the set on a particular feature, with the result that higher information indicates more usefulness [34]. By applying Information Gain after the feature extraction process, it is expected that the computation when building machine learning models is not too large so that the accuracy produced later can be optimized.

D. Chi-Square

Chi-Square is one of the feature extractions that measures the difference of data sharing by assuming features are independent of class values. It aims to determine the relationship in each attribute and then select features based on the highest results to form new features. The Chi-Square value is usually calculated from the confusion matrix [35], [36]. Using Chi-Square after the feature extraction process, it is expected that the computational burden in building machine learning models is not too high and the resulting accuracy can be more optimal.

E. Recursive Feature Elimination (RFE)

RFE is one of the cross-validation-based feature selections by recursively selecting features and arranging them like a ranking. In addition, RFE also tries to weed out collinearity and dependency in each feature. When all features in the dataset have been used, RFE will build a model by removing irrelevant features and will then use the remaining features [37].

3.5. Machine Learning

The deception detection model will concentrate on two machine learning namely, Random Forest and Support Vector Machine. The data used consists of numerical representations such as MFCC, Jitter, F0, and PLP that have been previously extracted and processed using openSMILE and Shennong in the python programming language. In order to make the machine learning model simpler and the extracted features more concise, the features will be reprocessed using various feature selection techniques. Then, the experimental results before and after feature selection are compared to find the features that are effective in eliminating the insignificant ones through the machine learning model.

A. Random Forest (RF)

Random Forest is one of the most commonly used machine learning techniques for classification and regression cases. Random Forest is an advanced development of the Decision Tree algorithm, which suffers from problems such as overfitting and the use of large data variances. Thus, Random Forest is expected to improve the accuracy of machine learning models without overfitting. The Random Forest approach combines all Decision Tree outputs into one output. Furthermore, based on the votes cast on the set of tree populations, the output results are randomly selected between asset attributes [38]. Fig. 6 is the process of Random Forest.

B. Support Vector Machine (SVM)

SVM is one of the machine learning methods that is often used in classification and regression cases such as Random Forest. The basic idea of SVM is to optimize the margin or distance between the hyperplane and the closest pattern and choose the best hyperplane perpendicular to the pattern. In addition, the advantage of SVM is that not all data is considered relevant during the iteration process. Later the contributing data is referred to as the support vector represented [39]. Fig. 7 is the process of SVM.
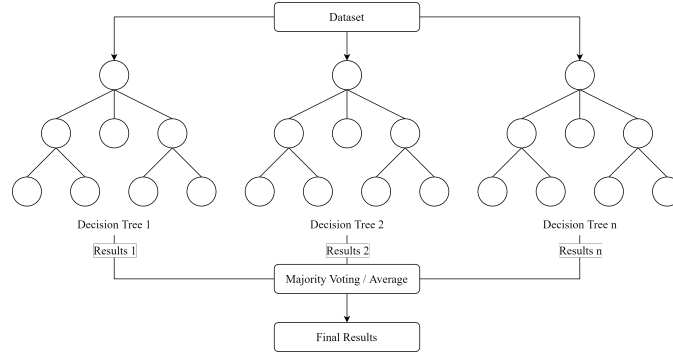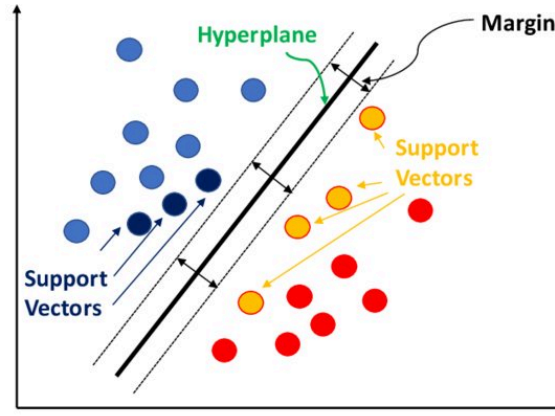
Fig. 6: Random Forest

Fig. 7: Support Vector Machine (SVM) [40]

3.6. Model Testing

The implementation of the model at this stage is to test the data that has been extracted from the openSMILE and Shennong libraries. The first stage of implementing the deception detection model is that the data will be cleaned first from unrelated features such as the file name of each audio data. Then, the data is trained with a ratio of 70:30 with a division of 70% for training data and 30% test data. In addition, hyperparameter tuning is also tested to determine which parameters are optimal for deception detection models.

3.7. Model Evaluation

There are two types of evaluation stages in deception detection, namely metric evaluation and statistical evaluation. Metric evaluation itself is generated from a model that has been tested with the aim of seeing how accurate the model is. Metric evaluation consists of $accuracy$, $precision$, $recall$, and $F1\text{-}score$. Accuracy is defined as the proportion of all predictions that are accurate predictions. In addition, the ratio of the number of positive data points classified to the total number of positive data points can also be used to determine precision. Then, recall is a term used to measure how well a model can classify relevant data. In addition to $Accuracy$, $Precision$, and $Recall$, $F1\text{-}score$ is a measure obtained by calculating the harmonic mean of $recall$ and $precision$ [41]. As for the statistical evaluation itself, the paired $t\text{-}test$ is used. The use of paired t-test is usually done if the p value which is the probability is less than the α value which is usually set at 0.05, then the null hypothesis can be rejected [42]. he purpose of using this $t\text{-}test$ is to determine whether or not there is a significant improvement in the evaluation before and after testing using feature selection [43]. The metric evaluation and statistical evaluation measures are defined in (1) to (5) where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. In the t-test formula, $\hat{d}$ is the sample mean of differences between paired observations, N is the sample size, and s_d^2 is the sample variance of the differences.
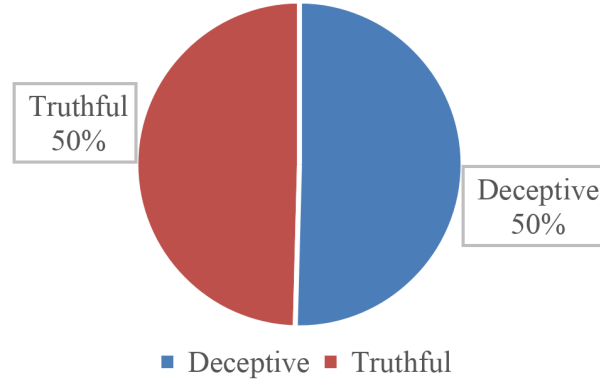
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Fig. 8: Distribution of Data Based on Labels Classed as Deceptive or Truthful

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\text{-}Score = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \tag{4}$$

$$T\text{-}Test = \frac{\hat{d}}{\sqrt{N^{-1}s_d^2}} \tag{5}$$

4. Results and Discussion

This research uses the English-based public dataset, Real-Life Trial from the University of Michigan-Deaborn. This data contains 121 videos with a division of 61 video data labeled deceptive or lying and 60 video data labeled truthful or honest with a percentage of 50:50. Because this dataset is still in video form, the data format is changed to audio data first before deception detection is carried out using the MoviePy library. After changing the data format, MFCC, Jitter, F0, and PLP features are extracted using 2 feature extraction libraries from the python programming language, namely, openSMILE and Shennong. For openSMILE itself uses the Interspeech 2016 Computational Paralinguistics Challenge (ComParE) feature set. The ComParE feature set produces 6373 static features by calculating various mathematical functions such as average, standard deviation, and so on the Low-Level Descriptor (LLD) contours including MFCC, Jitter, and F0 [44]. Then for Shennong itself extracts PLP as many as 14 coefficients. Then the two data are combined into one with a total of 6387 features. Fig. 8 is distribution of data based on label Deceptive and Truthful.

After extraction, this experiment will use 2 machine learning schemes, namely Random Forest (RF) and Support Vector Machine (SVM) with several feature selections such as Feature Importance, PCA, Information Gain, Chi-Square, and RFE. First, the data is preprocessed by removing irrelevant columns such as audio file names. Then, the data were initiated into a deception detection model with a data division of 70% for training data and 30% for testing data. The baseline method was tested with the hyperparameter tuning represented in Table 1. After that, another test was carried out with feature selection. At this stage, feature selection will produce an accuracy output based on the percentage of selected features from 5% to 100%. Fig. 9 is the result of the deception detection using Random Forest represented through a diagram.

In the experiment using random forest, there are seven experiments to compare which feature selection can improve the accuracy of the random forest-based deception detection model. In the first experiment, only feature importance was used. The accuracy before the feature importance feature selection in this experiment only got 72.97%, then the feature importance feature selection increased the accuracy to 83.78% with the best feature

Table 1: Hyperparameter tuning in machine learning.

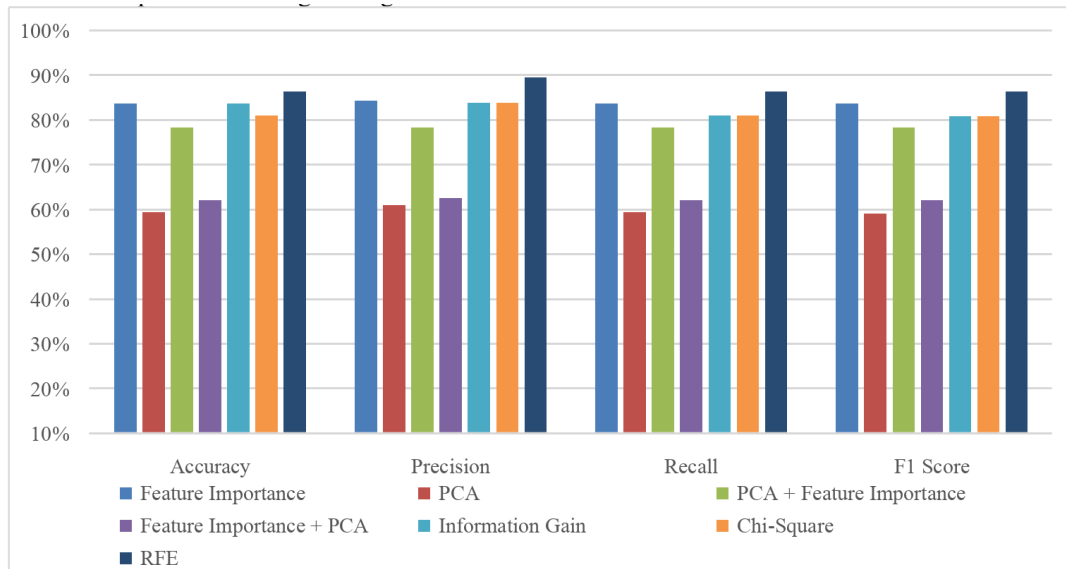| Machine Learning Model	Hyperparameter Tuning
Random Forest (RF)	max_depth = 4 n_estimators = 10 max_features = log2 min_samples_leaf = 2 min_samples_split = 5 random_state = 42
Support Vector Machine (SVM)	C = 0.9, random_state = 42

Fig. 9: Deception Detection Accuracy Results using a Random Forest

percentage at 50% with a total of 3193 features. Furthermore, in the second experiment, only Principal Component Analysis (PCA) was used. Before the dimension reduction using PCA, the accuracy of this experiment obtained the same results as before, namely 72.97%. However, when a reduction of 0.95 was carried out with 89 components, which means a reduction of 6297 features, there was a decrease in accuracy of 59.46%. Furthermore, in the third experiment, namely using a combination of PCA and feature importance, this experiment first performed PCA taking 89 components. Then, the experiment was carried out, and the results obtained were the same as the previous experiment, namely 59.46%. Then, the 89 components of the PCA reduction results were entered into the feature importance feature selection. After entering the feature importance feature selection, there was an increase in accuracy to 78.38% with the best feature percentage at 35% with 31 components. Furthermore, in the fourth experiment, the method used was actually the same, but only reversed the feature importance selection process first, then PCA was carried out. From this experiment, the feature importance produced the best accuracy of 81.08% with a percentage of 75% and 4789 features. After obtaining the best features of 4789 from the feature importance results, the best features were reduced using PCA with a reduction result of 86 components, which means that 4703 features were not used. From the PCA itself, there was a significant decrease in accuracy to 62.16%. Then, in the fifth experiment, we used the information gain feature selection. Before the feature selection was performed, testing was carried out using a baseline with an accuracy of only 72.97%. After feature selection, there was a fairly good increase in accuracy, which was 86.49% with the best percentage at 40% of the data totaling 2554. In addition to the use of information gain, in the sixth experiment using the Chi-Square, the accuracy was the same when tested at baseline. However, when feature selection was carried out, the accuracy increased to 81.08% at 80% of the data totaling 5108 features. In the last experiment using RFE, the baseline itself is also the same, which is 72.97%. However, when feature selection is performed using RFE, there is an increase in accuracy to 86.49%. This result is quite satisfactory when viewed in terms of accuracy. However, more features are taken, namely 2235 features.
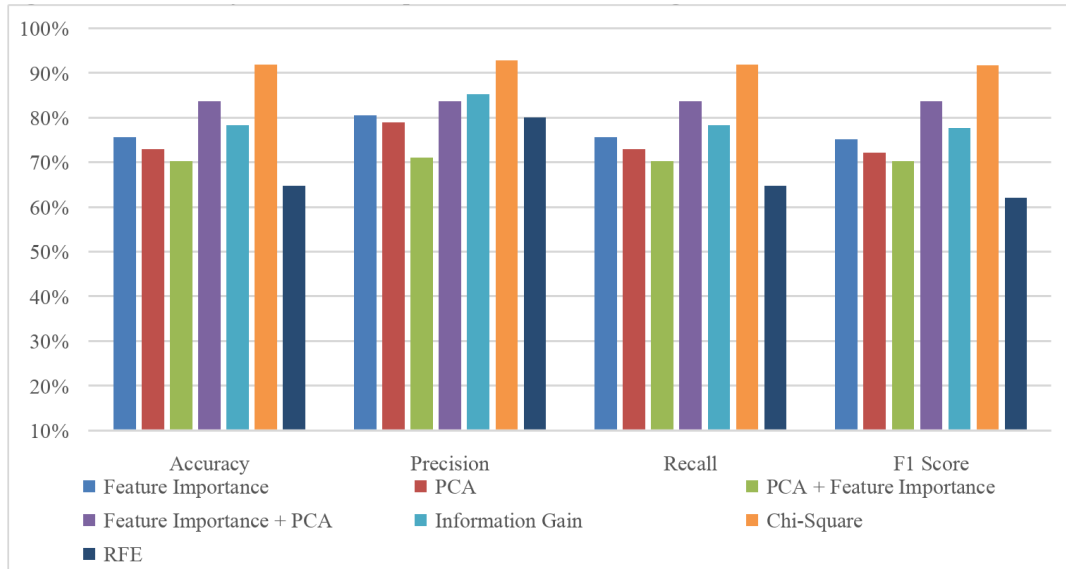
Fig. 10: Deception Detection Accuracy Results using a Random Forest

Besides using random forest as a deception detection model, this research also uses SVM as a comparison. Fig. 10 is a diagram of the accuracy results on deception detection when using SVM.

The use of SVM in this experiment has the same scheme as the experiment when using random forest, only the model and hyperparameter tuning used are different. In the first experiment using feature importance, the accuracy obtained before the feature importance feature selection in this experiment was 67.57%, which was then carried out with feature importance feature selection. There was an increase in accuracy to 75.68% with the best feature percentage at 35% with 2235 features. Furthermore, in the second experiment using PCA, an accuracy of 67.57% was obtained before the dimension reduction was carried out, but when the dimension reduction was applied, the accuracy of this experiment increased to 72.97%. Furthermore, the third experiment combined PCA and feature importance. importanteriment carried out PCA first taking 89 components from 6297 features. Then, the experiment was carried out, and the results obtained were the same as the previous experiment, namely 72.97%. Furthermore, from the 89 components of the PCA reduction results, they were entered into feature importance. After the experiment, there was a decrease in accuracy to 70.27% with the best feature percentage at 70% with 62 components. As for the fourth experiment, the opposite was true of the third experiment. From this experiment, the feature importance produced the best accuracy of 83.78% with a percentage of 5% and 319 features. After obtaining the best features of 319 from the feature importance results, the best features were reduced using PCA with a reduction result of 61 components, which means that 258 features were not used. After PCA is carried out from the feature importance results, the accuracy does not change, which is still at 83.78%. In the fifth experiment using information gain feature selection. Before feature selection, testing was carried out using a baseline with an accuracy of only 67.57%. After feature selection, there was a fairly good increase in accuracy of 78.38% with the best percentage at 35% with 2235 features. In the sixth experiment using the chi-square test, the accuracy was also the same when testing using a baseline. However, when feature selection was carried out, the accuracy increased to 91.89% at a percentage of 5% of the data totaling 319 features. In the last experiment using RFE, the baseline itself was also still the same, namely 67.57%. However, when feature selection using RFE was carried out, the accuracy decreased by 64.86% with the best feature percentage being 5% with 319 features. Using the RFE method in SVM can cause compatibility issues, which in turn results in decreased accuracy. however, fewer features are used when using random forest which is taken as 2235 features. The results of this experiment show that the deception detection model has improved after the feature selection process was carried out to reduce irrelevant features. In addition, when comparing the number of features selected between the two machine learning in this experiment, the random forest tended to use a medium to high number of features more often when selected. The SVM usually uses a smaller number of features than the results of the feature selection. However, when compared in terms of accuracy, SVM is

Table 2: Comparison number of features before and after feature selection and metrics evaluation.

Model	Feature Selection	Number of Feature Before Feature Selection	Number of Feature After Feature Selection	Accuracy	Precision	Recall	F1-Score
RF	FI	6387	3193	83.78%	84.32%	83.78%	83.81%
	PCA	6387	89	59.46%	61.10%	59.46%	59.16%
	PCA + FI	89	31	78.38%	78.38%	78.38%	78.38%
	FI + PCA	4789	86	62.16%	62.64%	62.16%	62.22%
	Information Gain	6387	638	83.78%	83.87%	81.08%	80.94%
	Chi-Square	6387	5108	81.08%	83.87%	81.08%	80.94%
	RFE	6387	2235	86.49%	89.56%	86.49%	86.39%
SVM	FI	6387	2235	75.68%	80.53%	75.68%	75.17%
	PCA	6387	89	72.97%	78.95%	72.97%	72.17%
	PCA + FI	89	62	70.27%	71.20%	70.27%	70.27%
	FI + PCA	321	61	83.78%	83.78%	83.78%	83.78%
	Information Gain	6387	2235	78.38%	85.30%	78.38%	77.73%
	Chi-Square	6387	319	91.89%	92.95%	91.89%	91.78%
	RFE	6387	319	64.86%	80.09%	64.86%	62.17%

far superior, even reaching 92% compared to random forest, which only gets its highest accuracy of 86%. Table 2 is the result of a comparison of the number of features before and after feature selection along with the evaluation.

Table 3: Comparison of accuracy prior to and post-feature selection and statistical analysis.

Model	Feature Selection	Accuracy Before Feature Selection	Accuracy After Feature Selection	Difference	T-Statistic	P-Value
RF	FI	72.97%	83.78%	10.81	−0.7774	0.4664
	PCA	72.97%	59.46%	−13.51		
	PCA + FI	59.46%	78.38%	18.92		
	FI + PCA	81.08%	62.16%	−18.92		
	Information Gain	72.97%	83.78%	10.81		
	Chi-Square	72.97%	81.08%	8.11		
	RFE	72.97%	86.49%	13.52		
SVM	FI	67.57%	75.68%	8.11	−1.7051	0.1390
	PCA	67.57%	72.97%	5,4		
	PCA + FI	72.97%	70.27%	−2,7		
	FI + PCA	83.78%	83.78%	0		
	Information Gain	67.57%	78.38%	10.81		
	Chi-Square	67.57%	91.89%	24.32		
	RFE	67.57%	64.86%	−2.71		

In addition to the use of metric evaluations such as accuracy, precision, recall, and f1-score. This study also uses statistical-based evaluations, namely paired t-tests. Paired t-tests are conducted to compare whether before and after selection there is a significant increase. Table 3 shows the results of a statistical evaluation of the accuracy before and after feature selection using paired t-tests. The t-test shows that before and after feature selection, there was no significant increase. This problem occurred in several experiments on the random forest and SVM models, which experienced a decrease in accuracy after feature selection with values close to those before feature selection. In the random forest model using PCA feature selection and a combination of feature importance and PCA, which previously obtained an accuracy of 72.97% and 81.08%, when feature selection was carried out there was a decrease

in accuracy to 59.46% and 62.16%. This discrepancy is also observed in SVM and random forest, particularly in experiments that employ a combination of PCA with feature importance and RFE, with accuracy rates dropping from 72.97% and 67.57% to 70.27% and 64.86% respectively.

However, when viewed on average in Table 3, the accuracy rate increases significantly. This increase in accuracy occurs in several feature selections such as RFE, information gain and PCA combined with feature importance, resulting in an accuracy of 86.49%, 83.78%, and 78.38% compared to before feature selection only getting 72.97% and 59.46% in the random forest model. Then, in experiments using chi-square feature selection, information gain and feature importance also experienced a significant increase, namely, 91.89%, 78.38%, and 75.68%, which before feature selection only got an accuracy of 67.57% using the SVM model.

5. Conclusions

This research uses the Real-Life Trial public dataset from the University of Michigan-Deaborn.Before the experiments, feature extraction was performed first, such as mel-frequency cepstral coefficient (MFCC), jitter, fundamental frequency (F0), and perceptual linear prediction (PLP).This experiment was conducted using random forest and support vector machine (SVM) with several feature selection methods namely feature importance, principal component analysis (PCA), information gain, chi-square, and recursive feature elimination (RFE).The use of feature selection is believed to improve machine learning models by reducing feature complexity such as irrelevant features. From this experiment, the use of PCA feature selection combined with feature importance and RFE in the random forest model increased the accuracy by 78.38% and 86.49%, respectively, which before feature selection only got an accuracy of 59.46% and 72.97%. From this accuracy, there was an increase of 18.92 and 13.52. Then, the use of PCA feature selection combined with feature importance and RFE succeeded in selecting 31 and 2235 features from the previous 6387 initial features. In addition, the chi-square and information gain also increased significantly when using the SVM. From the experiments conducted, chi-square and information gain get an accuracy of 91.89% and 78.38%, respectively, which before selection only get 67.57%. The increase in accuracy when chi-square and information gain have been done is 24.32 and 10.81. In addition to increasing accuracy, the use of feature selection such as chi-square and information gain also succeeded in selecting features to 319 and 2235 features from the initial 6387 features. This research requires further exploration such as the use of other feature selection methods or hyperparameter tuning, even machine learning or deep learning models in improving the deception detection model itself.

CRediT Authorship Contribution Statement

Muhammad M. Mafazy: Conceptualization, Software, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Supervision. **Chastine Fatichah:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition. **Anny Yuniarti:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The dataset was openly provided [https://public.websites.umich.edu/~zmohamed/resources.html].

Declaration of Generative AI and AI-assisted Technologies in The Writing Process

The authors used generative AI to improve the writing clarity of this paper. They reviewed and edited the AI-assisted content and take full responsibility for the final publication.

References

[1] A. Adha, "The Perception of Lying of Indonesians Living Abroad," *Kongr. Int. Masy. Linguist. Indones.*, 2022, doi: 10.51817/kimli.vi.7.

[2] J. Fan dan X. Shen, "New progress in the paradigm of elicited deception: Application of human-computer interaction in deception detection," in *2021 2nd International Conference on Information Science and Education (ICISE-IE)*, 2021, pp. 1558–1562, doi: 10.1109/icise-ie53922.2021.00345.

[3] T. Fornaciari dan M. Poesio, "Automatic deception detection in Italian court cases," *Artif. Intell. Law*, vol. 21, no. 3, 2013, doi: 10.1007/s10506-013-9140-4.

[4] V. Blikhar, O. Zaiats, N. Pavliuk, and N. Kalka, "Psychological and Legal Aspects of Verification and Detection of Lies during Polygraph Examination," *BRAIN. Broad Res. Artif. Intell. Neurosci.*, vol. 13, no. 1, 2022, doi: 10.18662/brain/13.1/284.

[5] Y. Li, J. Bian, and R. Song, "Video-based deception detection using wrapper-based feature selection," in *2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2024, pp. 1–5, doi: 10.1109/civemsa58715.2024.10586610.

[6] Z. K. Abdul dan A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," *IEEE Access*, vol. 10. 2022. doi: 10.1109/ACCESS.2022.3223444.

[7] D. Muttaqin dan S. Suyanto, "Speech Emotion Detection Using Mel-Frequency Cepstral Coefficient and Hidden Markov Model," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, 2020. doi: 10.1109/ISRITI51436.2020.9315433.

[8] K. Nishikawa, H. Kawano, R. Hirakawa, and Y. Nakatoh, "Analysis of Prosodic Features and Formant of Dementia Speech for Machine Learning," in *Proceedings - 2022 5th International Conference on Information and Computer Technologies, ICICT 2022*, 2022. doi: 10.1109/ICICT55905.2022.00037.

[9] W. Wang, X. Shen, H. Feng, B. Wang, and T. Yu, "Machine Learning based Deceptive Speech Detection," in *2nd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2023 - Proceedings*, 2023. doi: 10.1109/ICSCDS56580.2023.10104721.

[10] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Appl. Acoust.*, vol. 158, pp. 107020, 2020, doi: 10.1016/j.apacoust.2019.107020.

[11] K. Aktürk and A. S. Keçeli, "Deep operational audio-visual emotion recognition," *Neurocomputing*, vol. 588, p. 127713, 2024, doi: 10.1016/j.neucom.2024.127713.

[12] D. Doreswamy dan M. Nigus, "Feature Selection Methods for Household Food Insecurity Classification," in *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*, 2020. doi: 10.1109/ICCSEA49143.2020.9132945.

[13] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, "A Comparative Study of Model-Agnostic and Importance-Based Feature Selection Approaches," in *Proceedings - 2023 IEEE 5th International Conference on Cognitive Machine Intelligence, CogMI 2023*, 2023. doi: 10.1109/CogMI58952.2023.00020.

[14] M. U. Sen, V. Perez-Rosas, B. Yanikoglu, M. Abouelenien, M. Burzo, and R. Mihalcea, "Multimodal Deception Detection Using Real-Life Trial Data," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, 2022, doi: 10.1109/TAFFC.2020.3015684.

[15] C. Vens dan F. Costa, "Random forest based feature induction," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2011. doi: 10.1109/ICDM.2011.121.

[16] Q. T. Duong dan V. H. Do, "Development of Accent Recognition Systems for Vietnamese Speech," in *2021 24th Conference of the Oriental COCOSDA International Committee for the Co-Ordination and Standardisation of Speech Databases and Assessment Techniques, O-COCOSDA 2021*, 2021. doi: 10.1109/O-COCOSDA202152914.2021.9660512.

[17] I. M. Fadhil dan Y. Sibaroni, "Topic Classification in Indonesian-language Tweets using Fast-Text Feature Expansion with Support Vector Machine (SVM)," in *2022 International Conference on Data Science and Its Applications, ICoDSA 2022*, 2022. doi: 10.1109/ICoDSA55874.2022.9862899.

[18] H. Nasri, W. Ouarda, and A. M. Alimi, "ReLiDSS: Novel lie detection system from speech signal," in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2016. doi: 10.1109/AICCSA.2016.7945789.

[19] N. Srivastava and S. Dubey, "Deception detection using artificial neural network and support vector machine," *Biomedical Research*, vol. 29, no. 10, 2018, doi: 10.4066/biomedicalresearch.29-17-2882.

[20] H. Javaid, A. Dilawari, U. G. Khan, and B. Wajid, "EEG Guided Multimodal Lie Detection with Audio-Visual Cues," in *2nd IEEE International Conference on Artificial Intelligence, ICAI 2022*, 2022. doi: 10.1109/ICAI55435.2022.9773469.

[21] J. Xie, M. Zhu, and K. Hu, "Fusion-based speech emotion classification using two-stage feature selection," *Speech Communication*, vol. 152, p. 102955, 2023, doi: 10.1016/j.specom.2023.102955.

[22] F. Al-Dhaher, D. Y. Mohammed, M. Khalaf, K. Al-Karawi, M. Sarfraz, and M. M. Al Maathidi, "Real-Time Lie-Speech Determination Using Voice-Stress Technology," *Iraqi J. Comput. Sci. Math.*, vol. 5, no. 2, pp. 81–93, 2024, doi: 10.52866/ijcsm.2024.05.02.008.

[23] S. V Fernandes dan M. S. Ullah, "Use of machine learning for deception detection from spectral and cepstral features of speech signals," *IEEE Access*, vol. 9, pp. 78925–78935, 2021, doi: 10.1109/access.2021.3084200.

[24] X. Liu, J. Hancock, G. Zhang, R. Xu, D. Markowitz, and N. Bazarova, "Exploring Linguistic Features for Deception Detection in Unstructured Text," in *Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, 2012. doi: 10.1109/HICSS.2003.1173793.

[25] V. Venkata Krishna Reddy, R. Vijaya Kumar Reddy, M. Siva Krishna Munaga, B. Karnam, S. K. Maddila, and C. Sekhar Kolli, "Deep learning-based credit card fraud detection in federated learning," *Expert Systems with Applications*, vol. 255, p. 124493, 2024, doi: 10.1016/j.eswa.2024.124493.

[26] R. K. Gupta, A. Hassan, S. K. Majhi, N. Parveen, A. T. Zamani, R. Anitha, B. Ojha, A. K. Singh, and D. Muduli, "Enhanced framework for credit card fraud detection using robust feature selection and a stacking ensemble model approach," *Results in Engineering*, vol. 26, p. 105084, 2025, doi: 10.1016/j.rineng.2025.105084.

[27] L. Chen, W. Su, Y. Feng, M. Wu, J. She, and K. Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," *Information Sciences*, vol. 509, pp. 150–163, 2020, doi: 10.1016/j.ins.2019.09.005.

[28] J. T. Yang, G. M. Liu, and S. C. H. Huang, "Emotion Transformation Feature: Novel Feature for Deception Detection in Videos," in *Proceedings - International Conference on Image Processing, ICIP*, 2020. doi: 10.1109/ICIP40778.2020.9190846.

[29] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception Detection using Real-life Trial Data," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, New York, NY, USA: ACM, Nov 2015, pp. 59–66. doi: 10.1145/2818346.2820758.

[30] M. K. Camara, A. Postal, T. H. Maul, and G. H. Paetzold, "Can lies be faked? Comparing low-stakes and high-stakes deception video datasets from a Machine Learning perspective," *Expert Syst. Appl.*, vol. 249, pp. 123684, 2024, doi: 10.1016/j.eswa.2024.123684.

[31] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: What speech, music, and sound have in common," *Front. Psychol.*, vol. 4, no. MAY, 2013, doi: 10.3389/fpsyg.2013.00292.

[32] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, "Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method," *Symmetry (Basel).*, vol. 14, no. 6, 2022, doi: 10.3390/sym14061095.

[33] A. Razzaque and D. A. Badholia, "PCA based feature extraction and MPSO based feature selection for gene expression microarray medical data classification," *Measurement: Sensors*, vol. 31, p. 100945, 2024, doi: 10.1016/j.measen.2023.100945.

[34] Jupriyadi, S. Ahdan, I. L. Putra, A. Sucipto, A. Suhartanto, and E. A. Z. Hamidi, "A Ranking-Based Feature Selection for Indoor Positioning System," in *2023 IEEE 9th International Conference on Computing, Engineering and Design, ICCED 2023*, 2023. doi: 10.1109/ICCED60214.2023.10425728.

[35] M. Dhalaria dan E. Gandotra, "Android malware detection using chi-square feature selection and ensemble learning method," in *PDGC 2020 - 2020 6th International Conference on Parallel, Distributed and Grid Computing*, 2020. doi: 10.1109/PDGC50313.2020.9315818.

[36] I. Sumaiya Thaseen dan C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, 2017, doi: 10.1016/j.jksuci.2015.12.004.

[37] O. Kornyo, M. Asante, R. Opoku, K. Owusu-Agyemang, B. Tei Partey, E. K. Baah, and N. Boadu, "Botnet attacks classification in AMI networks with recursive feature elimination (RFE) and machine learning algorithms," *Computers & Security*, vol. 135, p. 103456, 2023, doi: 10.1016/j.cose.2023.103456.

[38] A. Singh, A. Singh, A. Aggarwal, and A. Chauhan, "Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2022*, 2022. doi: 10.1109/ICECCME55909.2022.9988588.

[39] L. Bergamin and F. Aiolli, "An investigation into creating counterfactual examples for non-linear Support Vector Machines," *Neurocomputing*, vol. 651, p. 130809, 2025, doi: 10.1016/j.neucom.2025.130809.

[40] O. N. Manjrekar dan M. P. Dudukovic, "Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique," *Chem. Eng. Sci. X*, vol. 2, 2019, doi: 10.1016/j.cesx.2019.100023.

[41] M. A. Nugraha, M. I. Mazdadi, A. Farmadi, Muliadi, and T. H. Saragih, "Penyeimbangan Kelas SMOTE dan Seleksi Fitur Ensemble Filter pada Support Vector Machine untuk Klasifikasi Penyakit Liver," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, 2023, doi: 10.25126/jtiik.1067234.

[42] A. R. M. Towfiqul Islam *etc.*, "Flood susceptibility modelling using advanced ensemble machine learning models," *Geosci. Front.*, vol. 12, no. 3, 2021, doi: 10.1016/j.gsf.2020.09.006.

[43] O. G. Atanda, W. Ismaila, A. O. Afolabi, O. A. Awodoye, A. S. Falohun, and J. P. Oguntoye, "Statistical Analysis of a Deep Learning Based Trimodal Biometric System Using Paired Sampling T-Test," in *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals, SEB-SDG 2023*, 2023, doi: 10.1109/SEB-SDG57117.2023.10124624.

[44] B. Schuller *etc.*, "The INTERSPEECH 2016 computational paralinguistics challenge: Deception, sincerity & native language," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016. doi: 10.21437/Interspeech.2016-129.