Javanese Gender Speech Recognition Using Deep Learning And Singular Value Decomposition

Kristiawan Nugroho[1,2], Edy Noersasongko[2], Purwanto[2], Muljono[2], Heru Agus Santoso[2]

[1]Computerized Accounting,[2] Faculty of Computer Science
[1]AMIK Jakarta Teknologi Cipta,[2]Dian Nuswantoro University
Semarang, Indonesia
[1]kristiawan1979@gmail.com,[2]edi-nur@dosen.dinus.ac.id,[2]purwanto@dsn.dinus.ac.id,
[2]muljono@dsn.dinus.ac.id, [2]heru.agus.santoso@dsn.dinus.ac.id

***Abstract*—Speech detection is an interesting research field. Research in speech recognition uses a variety of models that aim to improve accuracy, one of which is by using Deep Learning, but high dimensional data problems are one of the problems that cause a decrease in the quality of speech recognition accuracy. This paper discusses the gender voice recognition of Javanese people who are processed using the Mel Frequency Cepstral Coefficient (MFCC) extraction feature, then voice classification is done using the Deep Learning method combined with the Singular Value Decomposition (SVD) method in reducing the dimensions of the data produced. By using a 70% split ratio for training data and 30% for testing data the results of the research show that the Deep Learning method's accuracy is 97.78% higher than the Logistic Regression method of 95.56% and SVM of 93.33%. Speech recognition research shows that the Deep learning and SVD method can be used in performing speech recognition with a high degree of accuracy.**

Keywords-Recognition, Speech, Deep, Learning, MFCC, SVD

I. INTRODUCTION

Automatic Speech Recognition (ASR) is a research field that challenges and makes many researchers interested in this research field. ASR is the process of converting speech signals into word sequences using algorithms that are run through computer programs [1]. Several large companies such as Google, Apple, Microsoft and Amazon have developed a speech recognition algorithm system to achieve high accuracy and minimize error rates [2] Industry and researchers do not stop developing speech recognition technology because they want to develop a form of technology that makes humans feel comfortable when interacting with computers, robots and other machines [3].

Gender recognition is one of the research topics that is currently developing. Some researchers conduct research in gender recognition through image [4] and sound [5]. Research on gender speech recognition has been carried out, among others, by Sadek Ali using the First Fourier Transform (FFT) which achieved 80% accuracy in speech recognition[6]. Other studies on gender speech recognition were also conducted by Meena using fuzzy logic and neural networks[7]. Pahwa et all conducted research on gender recognition using SVM and neural networks resulting in an accuracy of gender voice recognition of 93.48%[8].

In the field of engineering, excessive dimensional data problems affect the efficiency of learning in machine learning and in the process of analyzing relationships between data or features [9]. Some algorithm models such as PCA (Principal Component Analysis) and SDD (Singular Data Decomposition) are used in solving these problems. This paper discusses research on Javanese gender voice recognition using the singular value decomposition method to reduce data with many dimensions and the use of deep learning methods in improving accuracy for performing gender speech recognition in Java language.

II. LITERATURE REVIEW

A. *Singular Value Decomposition(SVD)*

SVD is a numerical analysis technique used to diagonalize matrices and has been widely applied to overcome various mathematical problems [10]. SVD is the decomposition of the matrix into 3 matrices namely U, S and V where S is the singular value of the matrix [11]. The formula for the SVD equation can be written as follows:

$$\mathbf{A}_{nxp} = \mathbf{U}_{nxn}\,\mathbf{S}_{nxp}\,\mathbf{V}^{\mathbf{T}}{}_{pxp}$$

Where :

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_{\mathrm{nxn}}$$

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_{\mathrm{pxp}}\ \text{(i.e. U and V are orthogonal)}$$

978-1-7281-3832-9/19/$31.00 ©2019 IEEE

B. Deep Learning

Deep learning (DL) is one method that is often used in the field of machine learning. DL is a machine learning model based on the principle of Artificial Neural Network (ANN) that has been successfully used for information retrieval [12], image recognition, object tracking and language processing [13]. The form of deep learning is the development of artificial neural network methods that add hidden layers, Deep learning can be seen through Figure 1 as follows:

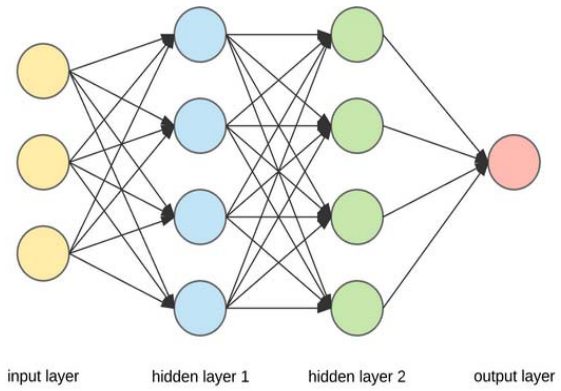

Fig. 1. Deep Learning Model.

The figure 1 shows that deep learning architecture begins with the input layer, in the deep learning method we can see the existence of several interconnected hidden layers between nodes with each other in processing the data entered so that the information will be through the output layer in this model. Various researches on gender speech recognition have been started many years ago. Sharma conducted research on speaker and gender identification. This study uses MFCC (Mel Frequency Cepstrum Coefficients) to perform extraction features and the radial base function method that successfully classifies with an accuracy rate of 96.11% [14]. The use of machine learning in detecting emotions and gender is also done using Adaboost with C4.5 reaching an accuracy rate of 93.3% [15]. Sedaaghi used Neural Network, SVM, KNN and GNN Classifier to recognize age and sex based on speech sounds [16], Use of the Neural Network method is also used by Seema Khanum in gender speech recognition in noisy environments, extraction features using MFCC, this method produces highest accuracy rate of 83.3% [17]. Other research on gender speech recognition was also conducted by Archana with extraction of features using MFCC and Artificial Neural Network and Support Vector Machine in classifying gender voices. This study resulted in a classification performance of 40% using ANN and 80% using SVM. In another study, Kumar used SVM to recognize emotions and types of gender voice. The extraction feature uses MFCC, using several datasets including the Berlin Emotional Database (BED) Reading, Leeds database, Belfast's database of research, producing accuracy between 80.4% and 84.5% [18].

III. METHODOLOGY

This study using machine learning with 4 stages of the model as described in Figure 2 as follows

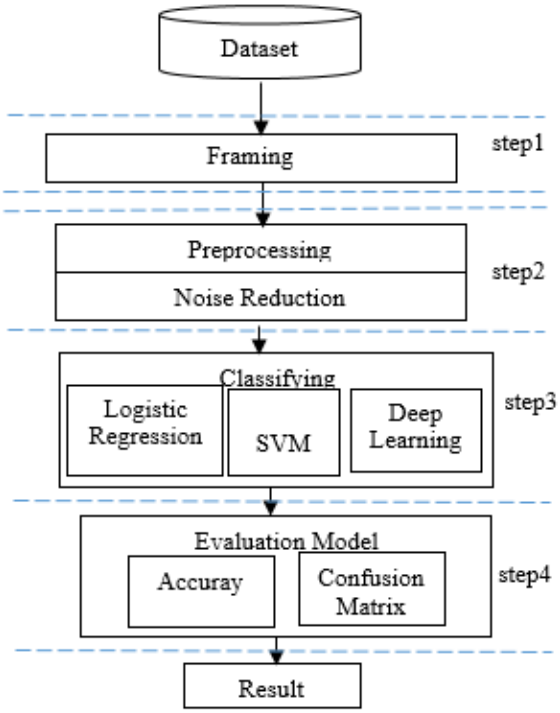

Fig. 2. Gender Recognition Proposed Method.

The figure 2 above shows machine learning processing using the source of the speech gender recognition dataset.

Step 1 : Shows the framing process where the sound of the gender sound recording is cut according to the predetermined time frame

Step 2 : Demonstrate the preprocessing process to process gender speech recognition datasets using noise reduction, which is to eliminate noise so that the dataset is cleaner and is expected to produce a high degree of accuracy

Step 3: Shows the classification process by comparing 3 methods namely Logistic Regression, Support Vector Machine (SVM) and Deep Learning. Comparison using these 3 methods serves to find the best accuracy method for Javanese gender speech recognition.

Step 4: An evaluation process from the results of the classification that has been done, this evaluation serves to calculate the level of accuracy and displays the results of confusion matrix which is a measurement of the results of machine learning classification.

IV. RESULT AND DISCUSSION

A. Dataset

This study used a gender-recorded gender speech dataset with male and female sex, there are 5 men and 5 women involved in this sound recording. Each gender says 3 words, namely "Eating", "Drinking", "Sleeping" each pronounced as many as 10 times. Sound recording is done using Adobe Audition software with Mono sound type, Sample Rate of 48000 Hz.

B.Preprocessing

Speech that have been recorded are then processed using Adobe Audition, the sound signals are then cut with the same duration of time which is equal to 31586. After the sound cutting is done and inserted in the excel 150 words will be processed in the next process. Preprocessing is a process carried out to process datasets by eliminating some parts that are not needed so that it is expected to produce a higher level of accuracy. This research used Noise Reduction facilities in Adobe Audition so that the sound produced will be clearer.

C. Fitur Extraction

The speech signal that has been preprocessed is then extracted by using the MFCC (Mel Frequency Cepstrum Coefficients) function in the MATLAB application. The results of the extraction feature are then labeled as male and female gender according to the recorded speech signal.

D. Gender Recognition & Evaluation

Gender classification is carried out with Rapidminer by using 3 classification methods namely Logistic Regression, SVM and Deep Learning with split validation. All methods used Singular Value Decomposition in reducing the number of high data dimensions. The research measurement results showed that Deep Learning obtained accuracy measurement results of 97.78%, classification error of 2.22%.

Gender speech classification using Support Vector Machine (SVM) obtained an accuracy rate of 93.33% and classification error of 6.67%. Gender voice classification using Logistic Regression obtained an accuracy rate of 95.56,% and classification error of 4.44%. The three models above produce same AUC (Area under the Curve of) of 0.998.

The results of the comparison of the 3 methods are shown in the figure 3 as below:

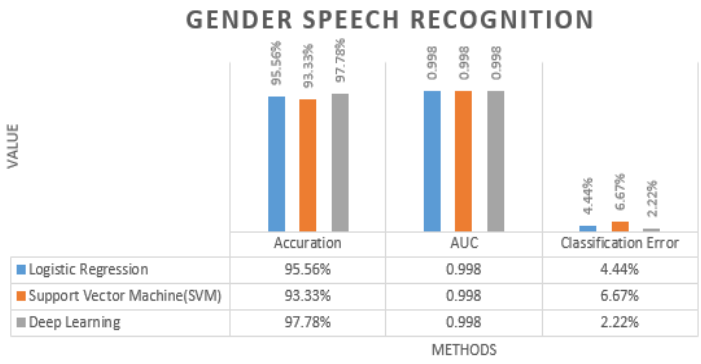

Fig. 3. Comparison of Gender Speech Recognition method.

E. Performance

The results of performance measurements in RapidMiner Studio used 3 methods displayed in the following tables :

1. Deep Learning (DL)

TABLE I. Performance Measurement DL

PerformanceVector:	accuracy: 97.78%	
ConfusionMatrix:		
True:	Woman	Man
Woman:	26	0
Man:	1	18
classification_error:	2.22 %	
AUC	0.998(positive class: Man)	

2. Support Vector Machine (SVM)

TABLE II. Performance Measurement SVM

PerformanceVector:	accuracy: 93.33%	
ConfusionMatrix:		
True:	Woman	Man
Woman:	27	3
Man:	0	15
classification_error:	6.67 %	
AUC	0.998(positive class: Man)	

3. Logistic Regression (LR)

TABLE III. Performance Measurement LR

PerformanceVector:	accuracy: 93.33%	
ConfusionMatrix:		
True:	Woman	Man
Woman:	25	0
Man:	2	18
classification_error:	6.67 %	
AUC	0.998(positive class: Man)	

V. CONCLUSION & FUTURE WORK

Research on Javanese gender speech recognition has been conducted on 10 speakers, 5 men and 5 women using Noise Reduction to eliminate noise in each gender voice data, the MFFC method was chosen as a good method to perform extraction features on 150 different gender sounds. The classification method used in the introduction of gender is Logistic Regression, SVM and Deep Learning resulting in different levels of accuracy, classification error and AUC where the Deep Learning method produces the highest accuracy rate of 97.78% when compared to other methods in gender speech recognition. In the next study, the Deep Learning method will be used to recognize the accent speech of various ethnic groups in Indonesia.

REFERENCES

[1] A. Kumar and V. Mittal, "Speech Recognition: A Complete Perspective," *Speech Recognit.*, vol. 7, no. 6, p. 6, 2019.

[2] G. Dharmale, D. D., and V. M., "Implementation of Efficient Speech Recognition System on Mobile Device for Hindi and English Language," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, 2019.

[3] A. N. Mon, W. Pa Pa, and Y. K. Thu, "Improving Myanmar Automatic Speech Recognition with Optimization of Convolutional Neural Network Parameters," *Int. J. Nat. Lang. Comput.*, vol. 7, no. 6, pp. 1–10, Dec. 2018.

[4] Z. Xie, Z. Guo, and C. Qian, "Palmprint gender classification by convolutional neural network," *IET Comput. Vis.*, vol. 12, no. 4, pp . 476–483, Jun. 2018.

[5] Y. Wang, J. Du, L.-R. Dai, and C.-H. Lee, "A Gender Mixture Detection Approach to Unsupervised Single-Channel Speech Separation Based on Deep Neural Networks," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1535–1546, Jul. 2017.

6] Md. Sadek Ali, "Gender Recognition System Using Speech Signal," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 2, no. 1, pp. 1–9, Feb. 2012.

[7] K. Meena, K. Subramaniam, M. Gomathy, and S. I. G. College, "Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network," vol. 10, no. 5, p. 9, 2013.

[8] Department of CSE, THE NORTHCAP University, Gurgaon, Haryana, India, A. Pahwa, and G. Aggarwal, "Speech Feature Extraction for Gender Recognition," *Int. J. Image Graph. Signal Process.*, vol. 8, no. 9, pp. 17–25, Sep. 2016.

[9] Z. Cheng and Z. Lu, "A Novel Efficient Feature Dimensionality Reduction Method and Its Application in Engineering," *Complexity*, vol. 2018, pp. 1–14, Oct. 2018.

[10] G. Zhang, W. Zou, X. Zhang, X. Hu, and Y. Zhao, "Singular value decomposition based sample diversity and adaptive weighted fusion for face recognition," *Digit. Signal Process.*, vol. 62, pp. 150–156, Mar. 2017.

[11] H. B. Razafindradina, P. A. Randriamitantsoa, and N. R. Razafindrakoto, "Image Compression with SVD : A New Quality Metric Based On Energy Ratio," vol. 5, no. 6, p. 6, 2016.

[12] S. G. Santur and Y. Santur, "Knowledge Mining Approach For Healthy Monitoring From Pregnancy Data With Big Volumes," *Int. J. Intell. Syst. Appl. Eng.*, vol. 4, no. Special Issue-1, pp. 141–145, Dec. 2016.

[13] Y. Santur, M. Karaköse, and E. Akın, "Condition Monitoring Approach Using 3D- Modelling of Railway Tracks With Laser Cameras," p. 5.

[14] S. Sharma, A. Shukla, and P. Mishra, "Speaker and Gender Identification on Indian Languages using Multilingual Speech," vol. 1, no. 4, p. 4.

[15] D. Kaur, "Machine Learning Based Gender Recognition And Emotion Detection," *Int. J. Eng. Sci.*, vol. 7, no. 2, p. 6.

[16] M. H. Sedaaghi, "A Comparative Study of Gender and Age Classification in Speech Signals," *Electron. Eng.*, vol. 5, no. 1, p. 12, 2009.

[17] S. Khanum, "Speech based Gender Identification using Feed Forward Neural Networks," *Int. J. Comput. Appl.*, p. 4.

[18] S. S. Kumar and T. RangaBabu, "Emotion and Gender Recognition of Speech Signals Using SVM," vol. 4, no. 3, p. 10, 2015.