Journal of Soft Computing Exploration
Homepage: htttps://shmpublisher.com/index.php/joscex
p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

Sound detection of gamelan musical instruments using teachable machine

Yessi Yunitasari[1*], Moch Yusuf Asyhari [2], Inung Diah Kurniawati [3], Latjuba Sofyana STT[4]
[1,,2,3,4] Department of Informatic Engineering, Universitas PGRI Madiun, Indonesia

Article Info

Article history:

Received May 15, 2025
Revised May 20, 2025
Accepted June 18, 2025

Keywords:

Sound detection
Gamelan
Teachable machine
CNN
Classification

ABSTRACT

Gamelan is an instrument of musical expression that has an aesthetic function related to social, moral, and spiritual values. Gamelan consists of a variety of musical instruments that have a unique sound. In this study, the sound detection of nine gamelan musical instruments was carried out using a teachable machine. The gamelan musical instruments detected included gong, kenong, saron, bonang, gambang, kendang, flute, siter, and rebab. The algorithm used is CNN. The CNN algorithm has a fairly good performance for the sound detection process. The test results of the built model show an "acc" value of 25 ranging from 0.99 to 1, which indicates that the model achieves an accuracy rate of 99% to 100% on the training dataset. At the same time, "test accuracy" refers to a measure of the model's effectiveness in predicting data it has not encountered during training. The "test accuracy" score varied from 0. 83, which shows that the validation data has an accuracy of 83%.

Corresponding Author:

Yessi Yunitasari,
Department of Informatic Engineering,
Universitas PGRI Madiun,
Indonesia
Email: yessi@unipma.ac.id
https://doi.org/10.52465/joscex.v6i2.576

1. INTRODUCTION

Gamelan is one of the traditional musical instruments from the island of Java. A set of gamelan consists of various musical instruments that are played in unison. Because the combination of the sound of one gamelan and the sound of the other gamelan illustrates harmony in community life. But as time goes by, there are fewer and fewer gamelan enthusiasts and gamelan makers are also increasingly difficult to find. The development of Javanese gamelan was initially introduced by the wali Songo to help spread Islam in the archipelago, especially the island of Java. The guardians used gamelan as an attraction so that the public would want to enter Islam and come to the Mosque to perform worship [1].

The overlapping frequency phenomenon in gamelan performance presents a complex challenge in automatic instrument identification. When gamelan instruments are played simultaneously, their harmonic spectra often intersect, which can disrupt the accuracy of detection systems. Therefore, advanced computational models are needed to recognize and classify these sounds correctly.

Artificial intelligence technology that focuses on using data and algorithms to mimic the way humans learn, one of these learning models is computational performance using deep learning techniques. Deep learning, particularly through the use of Convolutional Neural Networks (CNNs), has demonstrated strong

performance in complex signal processing, especially in the field of speech recognition. For instance, research [2] and [3] successfully implemented neural networks and MFCC feature extraction to recognize speech with high accuracy, such as the Darija and Kambera dialects, achieving results up to 99.6%. These approaches leverage spectrogram-like inputs and layered convolutional filters to capture nuanced patterns in audio signals.

However, these models have mainly focused on human speech and dialects, with limited application to non-verbal or instrumental sounds such as gamelan music. Furthermore, gamelan differs fundamentally from speech in terms of timbre, pitch layering, and rhythmic complexity, highlighting a research gap that remains unexplored in current literature. Survey results related to Automatic Speech Recognition using Advanced Deep Learning Approaches namely Advances in deep learning (DL) towards automatic speech recognition (ASR). ASR relies on large training datasets, and requires substantial computational and storage resources. Enabling adaptive systems can improve the performance of ASR in dynamic environments [4].

The methods discussed in [5]–[7] emphasize the potential of CNNs not only in speech but also in animal sound recognition, such as elephant vocalizations. These studies used dual-layered convolutional systems to extract deep features and achieved accuracy rates up to 96.2%. The common thread among these works is the use of CNNs as a generalized audio classifier; however, none have addressed the unique acoustic patterns in gamelan music.

Quoc Bao Diep et al. [7] proposed three CNN-based approaches (1D-CNN, 2DS-CNN, and 2DM-CNN) and achieved performance exceeding standard models like GoogLeNet and AlexNet with accuracies up to 99.76%. This demonstrates that CNN architectures are scalable and adaptable across domains. However, applying them to gamelan instruments, which often feature overlapping percussive and melodic elements, requires careful consideration of feature extraction techniques and temporal dynamics. Convolutional neural networks (CNNs) are deep learning architectures that are often used to solve a variety of problems. Convolutional neural network (CNN) is a method that can help solve the problem of large amounts of data, and the CNN method can help solve the demands of user identification problems [8].

Several researchers have explored the application of deep learning, particularly Convolutional Neural Networks (CNNs), in audio classification and traditional instrument recognition. For instance, the study by Muljono et al. employed a combination of Mel-Frequency Cepstral Coefficients (MFCCs) and CNNs to classify tempo in kendhang, a key instrument in the gamelan ensemble, achieving an accuracy of up to 97% [9]. Another study by Hermawan et al. applied Short-Time Fourier Transform (STFT) and multilayer perceptron (MLP) models for the transcription of musical notes from the Demung instrument in gamelan, demonstrating the effectiveness of deep learning for traditional sound processing [10].

The other research showed that CNNs trained with log-Mel or MFCC spectrogram inputs could hierarchically learn audio features for instrument recognition with high accuracy [11]. Their work emphasized the interpretability of CNN layers, showing that deeper layers captured timbral characteristics specific to instruments, thereby improving classification performance. Giri and Radhitya further confirmed the strength of CNNs in this domain by experimenting with mel-spectrogram and MFCC inputs for a dataset of various musical instruments, including traditional ones. Their model achieved high accuracy and robustness, suggesting its potential use for instrument classification tasks such as those in gamelan orchestras [12].

In our study, we applied a CNN-based deep learning approach using Teachable Machine combined with MFCC feature extraction to recognize many different instruments in a gamelan ensemble such as rebab, gender penerus, gender barung, bonang barung, saron barung ,slenthem, demung, saron penerus, gambang, clempung, siter, kenong, kempul, engkuk-kemong, kemanak, suling, kethuk-kempyang, gong suwukan, gong ageng, kendang. We performed experiments across 10, 15, 20, and 25 epochs to achieved the highest accuracy on the test. This confirms that CNN models, even when trained with moderate resources, can capture the essential acoustic signatures of gamelan instruments or not.

Despite the promising results of CNN and MFCC in general audio classification tasks, their application in gamelan sound recognition remains limited and underexplored. This study fills that gap by exploring the intersection of traditional gamelan instrumentation and modern deep learning classification models, ultimately contributing toward the digital preservation and modernization of Indonesian cultural heritage.

2. METHOD

In this study, a teachable machine with a CNN algorithm was used to classify the types of sounds in a sample of gamelan music. From the gamelan music sample, the sound of 9 musical instruments will be

detected. The musical instruments detected include: gong, kenong, saron, bonang, gambang, kendang, flute, siter, and rebab. The workflow of the gamelan sound detection process can be seen in the image below:

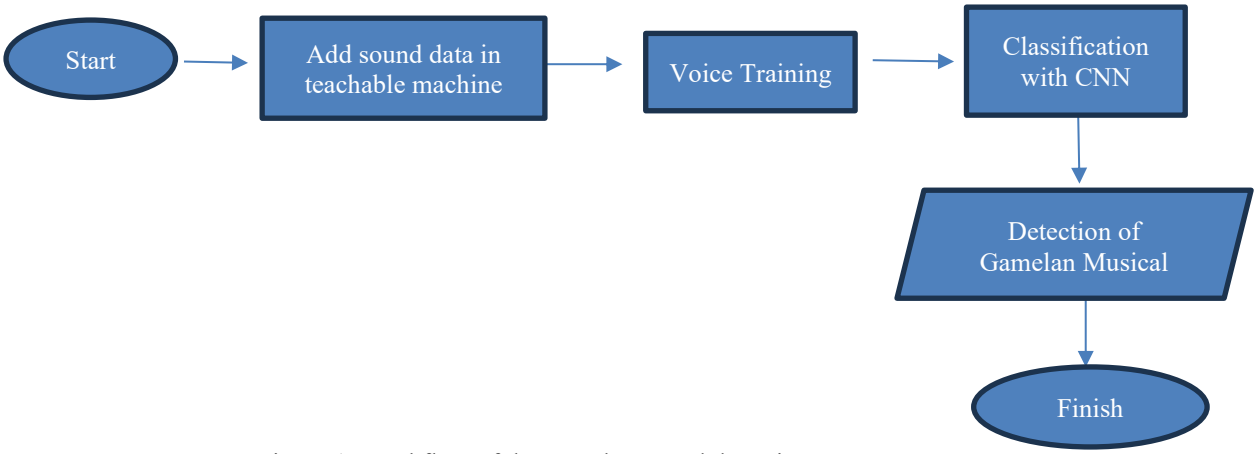

Figure 1. Workflow of the gamelan sound detection process

1. Create a new Project

The first step is to create a new project on the teachable machine. Teachable machine can be used for 3 features. First, the image feature, the sound feature, and the pose feature. Choose for the sound project, here is the view for the sound feature on the tachable machine.

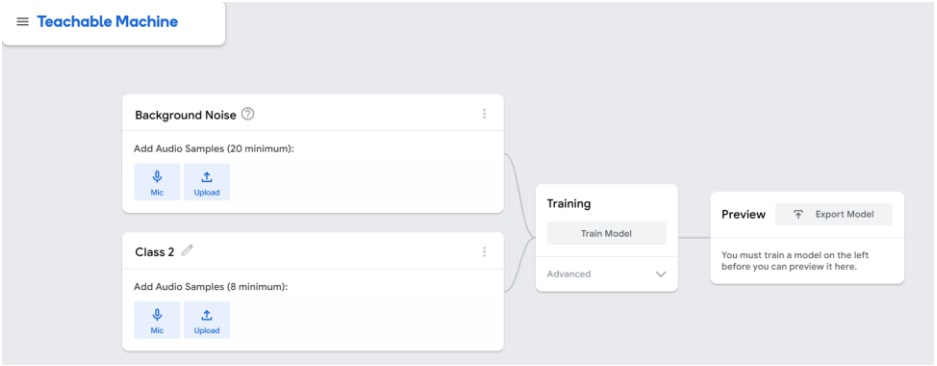

Figure 2. Tachable machine sound fiture.

2. Add Voice Data in Teachable Machine

The second step we need to record background noise, background noise needs to be recorded to know the noise in our surroundings. We can record the noise for 20 seconds and extract the sample. The more sample data provided, the better. At the time of noise recording we can also set the recording duration and delay. The duration and delay settings can be seen in Figure 3.

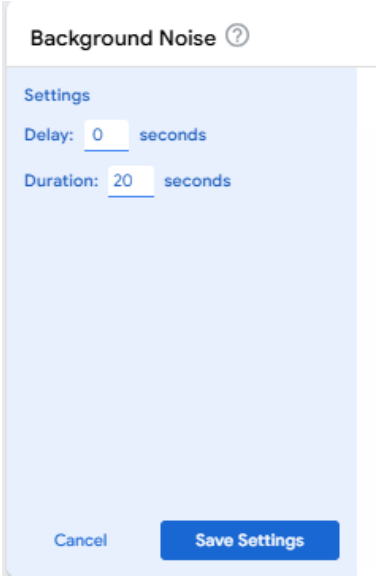

Figure 3. Setting delay and recording duration

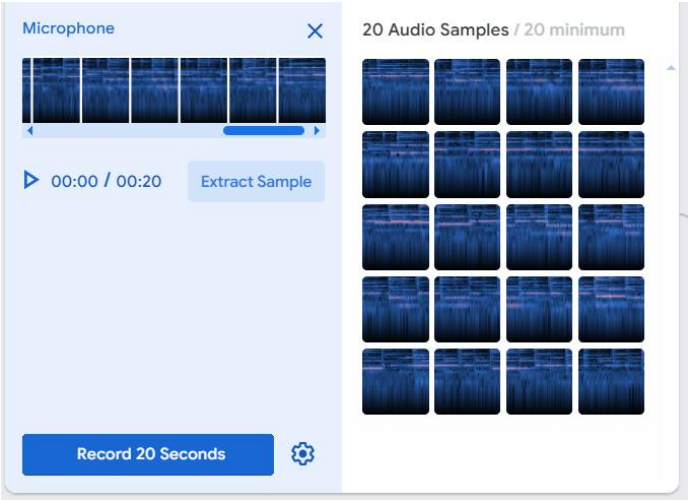

Figure 4. Add sample noise

The microphone used to record sound is miniso headphone. The number of sample data used is 216 data with 9 classes. The recording environment is carried out in a medium-sized room, with minimal echo or reverberation. After recording the noise we need to add a sound sample that will later be detected. In this study, 9 Javanese gamelan musical instruments were used to be detected. Musical instrument sound samples are recorded with the help of a microphone connected to a laptop and stored directly in a tachable machine. After we get a sample of the sound of gamelan music consisting of various kinds of musical instruments. We will add the sound samples that we get to the teachable machine.

The sound sample was sampled for 20 seconds and then extracted into different audio samples. If the extracted sound sample is not suitable, we can repeat the steps to readd the sound sample. Samples of gamelan musical instruments used include gong, kenong, saron, bonang, gambang, kendang, flute, siter, and rebab. Sound sampling is done by following the standard recording duration of 2 s per instrument sample, with the option to re-record if needed. An example of a sample of gamelan musical instruments can be seen in figure 5.

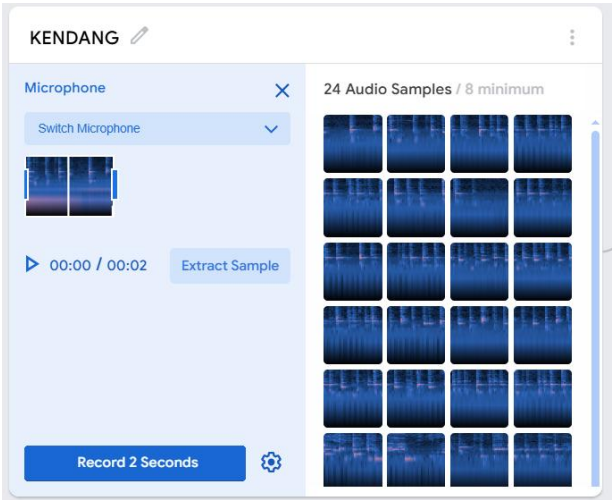

Figure 5. Sample kendang musical instrument

3. Voice Training and Classification with CNN

Training and classification were carried out using the TensorFlow framework which has been equipped with the CNN algorithm. The TensorFlow framework and CNN algorithm are used to measure the sound recognition performance of gamelan musical instruments. The design of the CNN model architecture was carried out to be able to recognize the sound patterns that have been recorded in the dataset and to be able to identify what is being spoken. The model was designed using the tensorflow framework that functions as a backend engine and a hard library that functions as a high-level neural networks API. The process of forming a neural network model in the creation of speech recognition is carried out by trying values and models until the model made is considered to be quite good in terms of accuracy and loss values that are not too high [13]. CNN leverage the spatial feature hierarchy of the audio data to be processed. While TensorFlow.js is an open-source machine learning library. It is based on the TensorFlow.js library written in Python. Learning knowledge to help learn something different but similar is called transfer learning. Machine learning models that have been drilled like images or audio data. Traditional Convolutional Neural Network (CNN) architectures can see how transfer learning can leverage these trained networks to learn new objects [14].

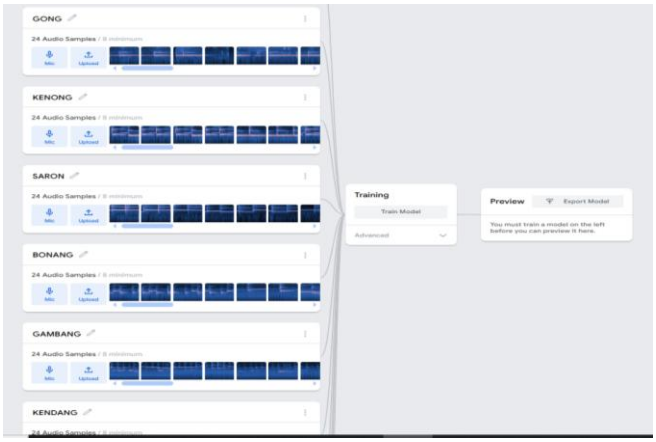

Figure 6. Training process

We can start the Training process by pressing the train model button. Epoch settings can be done by clicking advance. We can try more than 1 epoch to get the best performance results. In the research, the detection of gamelan musical instruments was carried out 4 epochs.

4. Detection Music Gamelan

Gamelan consists of two adjustment systems, namely the barrel of slendro and pelog, with the ricikan gamelan consisting of: rebab, gender penerus, gender barung, bonang barung, bonang barung, saron barung ,slenthem, demung, saron penerus, gambang, clempung, siter, kenong, kempul, engkuk-kemong, kemanak, suling, kethuk -kempyang, gong suwukan, gong ageng, kendang.[15]. The gamelan music detection display on the teachable machine we can see the spectogram. A spectrogram is a visual representation of a sound that

shows the frequency and timing distribution of that sound. Spectrograms are created using a Fourier Transform that converts sound signals from the time domain to the frequency domain, then displayed in the form of an amplitude graph of frequency and time [16]. In addition, we can set the overlap factor. This overlap factor determines how often the last second of audio is tested against the model you created. In this study, the overlap factor used is 0.5 which means that audio will be grouped every half second. Meanwhile, if we choose the overlap value of the factor of 0, the audio data will be compressed every 1 second. The gong sound sample when we tested showed a percentage of 37% more dominant among other models. Another sample when we try the sitter sound shows that the percentage is 72% more dominant among other gamelan musical instruments. This shows that the model can detect well. In a speech recognition system, there is an input layer to receive the original speech signal. While the non-linear activation function can change the input data, and the hidden layer plays a role in extracting features based on the data. The output layer is responsible for mapping these features to specific speech units.[17].

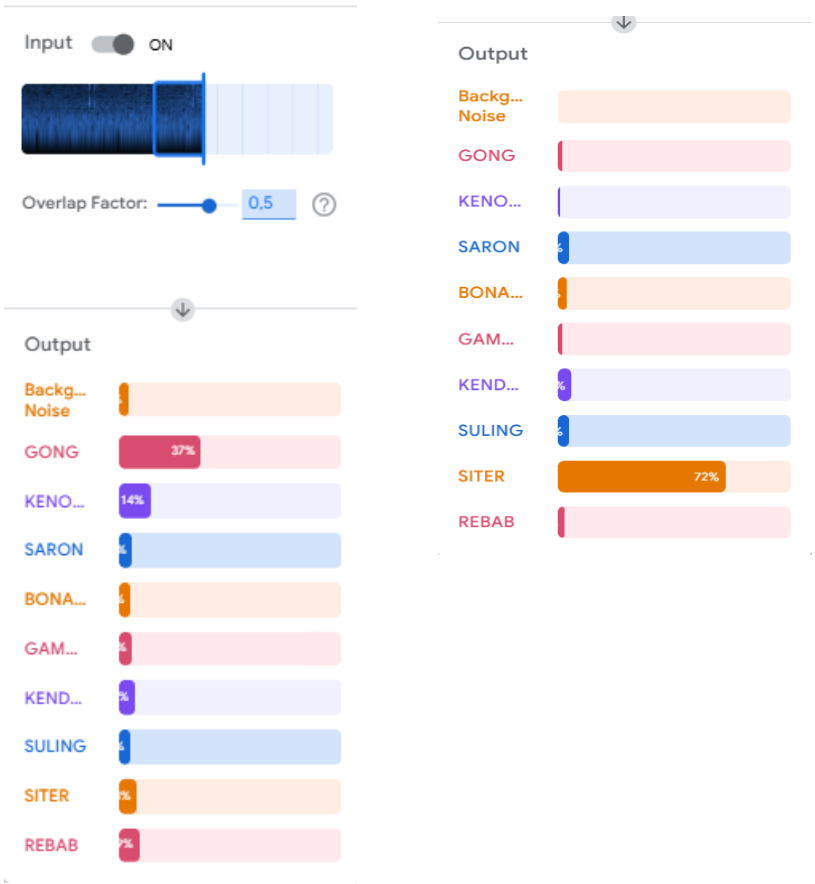

Figure 7. Trial of gong and siter musical instruments.

After we test the model we can export the model. We can also export our trained templates for our projects: websites, applications, etc. We can download our trained template or host it online for free from there and use it anywhere Javascript is run.

Figure 8. Export model

3. RESULTS AND DISCUSSIONS

The results of the model built we can see the visual display of the accuracy graph per epoch in the screenshot on the Teachable Machine site using the CNN algorithm. CNN is a type of artificial neural network that is typically used to identify objects [18]. Convolutional Neural Network (CNN) is used as a place to process data in two-dimensional form (multi layer perceptron), namely voice data and image data. CNN as one form of pattern recognition that requires deep learning to detect it [19]. Model training is the process of training to recognize objects and classify them according to their class [20]. The "acc" metric is used to measure accuracy in model training and shows the model's ability to predict training data. In this study, several experiments were carried out to change the epoch value to determine the results of the model's performance. The results of the experiment used different epoch values along with the loss and test loss values obtained. In the use of speech recognition through Teachable Machine, the term "accuracy per epoch" refers to a measure that shows the percentage of data that is successfully classified correctly by the model during one training cycle (epoch). The results of the comparison of epoch values, accuracy, test accuracy, loss values and test loss can be seen in Table 1.

Table 1 Results of the epoch experiment

Epoch	Acc	Test Acc	Loss	Test Loss	Precision	Recall	F1-Score
10	0,9	0,69	0,39	1,08	0,70	0,87	0,61
15	0,94	0,74	0,26	0,87	0,75	0,79	0,53
20	0,95	0,81	0,2	0,56	0,65	0,67	0,72
25	0,99	0,83	0,14	0,61	0,79	0,76	0.80

The “acc” value for epoch 25 ranges from 0.99 to 1, meaning the model has an accuracy of between 99% and 100% against the training data. Meanwhile, “test acc” is the validation accuracy that measures the model’s ability to predict test data that is not used in training. The “test acc” value ranges from 0.83, indicating an accuracy of between 83% against the validation data. The results of the accuracy and test accuracy graphs for epoch 25 can be seen in Figure 9.

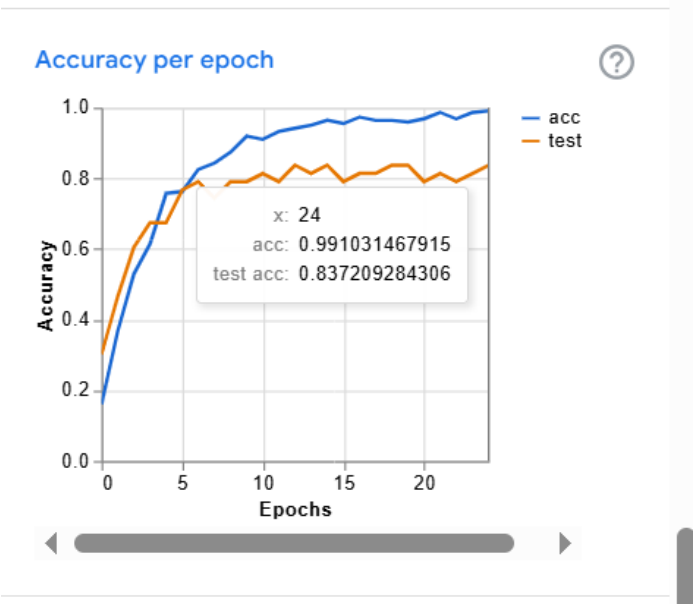

Figure 9. Accuray and test accuracy results for epoch 25

The test results for epoch 25 show that the loss value obtained is 0.14 while the test loss value is 0.61. The graph results of the Loss and Test Loss Results for epoch 25 can be seen in Figure 6. On the other hand, “loss per epoch” shows the level of model prediction error against training data in one epoch. The target of model training is to improve the accuracy per epoch while reducing the loss per epoch, which can be achieved by adjusting model parameters such as learning rate or changing the model structure so that the model can be more effective in learning and classifying sounds according to the desired application goals.

Figure 10. Loss and test loss results for epoch 25

A graph of loss change per epoch in a screenshot on the Teachable Machine website, which illustrates the model's ability to predict the entire dataset in each training iteration. The "loss" metric measures the difference the model's prediction from the actual value. The lower the loss value, the better the model in making predictions. In addition, there is a test loss metric, which measures model errors in each epoch's test or validation data. This test loss is important to evaluate the model's ability to apply learning from training data to data that has never been used before. The results of the test loss, with a range of values between 1.08 to 0.61

on the validation data, showed a consistent decline. The Epoch Test Result Bar Chart which contains the comparison of results for accuracy, accuracy test, loss and test loss can be seen in Figure 11.

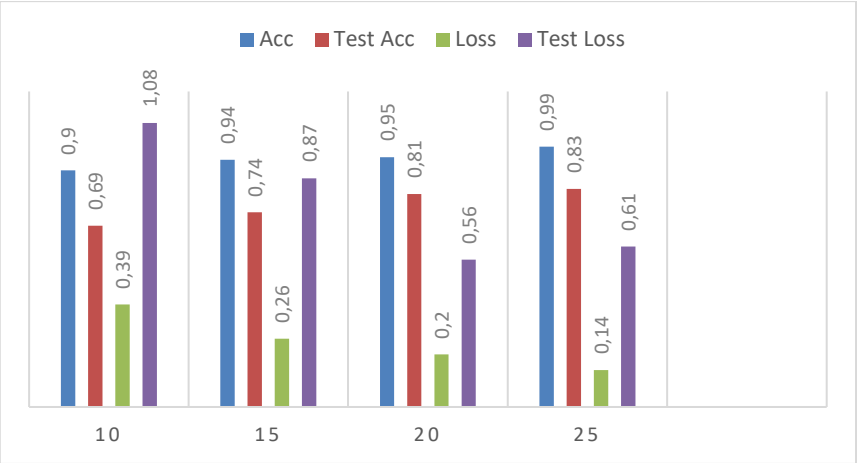

Figure 11. Epoch test result bar chart

From the graph we can see that the model results are experiencing overfitting. This shows that the model focuses too much on the details of the training data and cannot generalize well. Overfitting usually occurs when a model has too many parameters or features. Such complex models tend to capture small details, including noise, in the training data. As a result, even though the model may show high accuracy when tested with training data, its performance will decline when faced with new data that it has never seen before.

4. CONCLUSION

In conclusion, we are satisfied with the final results of our Teachable Machine model to detect gamelan musical instrument sounds based on the TensorFlow deep-learning model. Although this model is not perfect, the model built can fully operate to detect 9 gamelan musical instruments tested. This model can be used easily. The limitation of this model is that the model is still very simple and can only detect 9 types of musical instruments. The model that was built is still overfitting, it is hoped that in the future it can be improved so that it becomes good fitting. So in the future, we will try to apply this model to other case studies and can be refined to detect gamelan musical instruments with more types of musical instruments. And can be implemented into a mobile application with a better and more attractive user interface.

REFERENCES

[1] E. Ariyanto and F. S. Hananto, "Identifikasi Dan Aplikasi Pengenalan Spektrum Bunyi Gamelan Menggunakan Jaringan Syaraf Tiruan Pada Matlab," *J. Neutrino*, vol. 7, no. 1, p. 7, 2014, doi: 10.18860/neu.v7i1.2633.

[2] M. Jebbar, A. Maizate, and R. A. Abdelouahid, "Moroccan's Arabic Speech Training And Deploying Machine Learning Models with Teachable Machine," *Procedia Comput. Sci.*, vol. 203, pp. 801–806, 2022, doi: 10.1016/j.procs.2022.07.120.

[3] E. A. U. Malahina, "Teachable Machine: Deteksi Dialek Sumba Timur (Kambera) Menggunakan Layanan Open Source," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 12, no. 4, pp. 280–286, 2023, doi: 10.22146/jnteti.v12i4.8174.

[4] H. Kheddar, M. Hemis, and Y. Himeur, "Automatic speech recognition using advanced deep learning approaches: A survey," *Inf. Fusion*, vol. 109, 2024, doi: 10.1016/j.inffus.2024.102422.

[5] D. T. Susetianingtias and E. Patriya, "Identifikasi Fitur Suara Menggunakan Model Convolutional Neural Network (CNN) pada Speech-to-Text (STT)," vol. 4, no. 3, pp. 809–820, 2024.

[6] T. Thomas Leonid and R. Jayaparvathy, "Classification of Elephant Sounds Using Parallel Convolutional Neural Network," *Intell. Autom. Soft Comput.*, vol. 32, no. 3, pp. 1415–1426, 2022, doi: 10.32604/IASC.2022.021939.

[7] Q. B. Diep, H. Y. Phan, and T. C. Truong, "Crossmixed convolutional neural network for digital speech recognition," *PLoS One*, vol. 19, no. 4, pp. 1–22, 2024, doi: 10.1371/journal.pone.0302394.

[8] W. Ibrahim, H. Candra, and H. Isyanto, "Voice Recognition Security Reliability Analysis Using Deep Learning Convolutional Neural Network Algorithm," *J. Electr. Technol. UMY*, vol. 6, no. 1, pp. 1–11, 2022, doi: 10.18196/jet.v6i1.14281.

[9] T. S. Prihartini and P. N. Andono, "Deteksi Tepi Dengan Metode Laplacian of Gaussian Pada Citra Daun Tanaman Kopi," pp. 1–5, 2017, [Online]. Available: http://eprints.dinus.ac.id/15312/1/jurnal_15354.pdf

[10] A. R. Hermawan, "276-1002-1-Pb," vol. 6, no. 2, pp. 0–6, 2022.

[11] R. Chen, A. Ghobakhlou, and A. Narayanan, "Interpreting CNN models for musical instrument recognition using multi-spectrogram heatmap analysis: a preliminary study," *Front. Artif. Intell.*, vol. 7, no. December, pp. 1–14, 2024, doi: 10.3389/frai.2024.1499913.

[12] G. A. V. M. Giri and M. L. Radhitya, "Musical Instrument Classification using Audio Features and Convolutional Neural Network," *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 226–234, 2024, doi: 10.30871/jaic.v8i1.8058.

[13] V. Chhipa and R. Poonia, "Teachable machine : A web based machine learning tool for user voice biometric authentication system," *J. Discret. Math. Sci. Cryptogr.*, vol. 27, no. 4, pp. 1345–1355, 2024, doi: 10.47974/JDMSC-1989.

[14] E. A. U. Malahina, R. P. Hadjon, and F. Y. Bisilisin, "Teachable Machine: Real-Time Attendance of Students Based on Open Source System," *IJICS (International J. Informatics Comput. Sci.*, vol. 6, no. 3, p. 140, 2022, doi: 10.30865/ijics.v6i3.4928.

[15] K. Hastuti and A. M. Syarifzh, "Identifikasi Fitur Melodi Dalam Musik Gamelan Berdasarkan Hubungan Asosiasi Antar-

Notasi," *Semin. Nas. Sist. Inf. Indones.*, pp. 47–54, 2016.

[16] I. M. Nurrizqy, B. H. Prasetio, and R. R. Mardi Putri, "Sistem Kontrol Perangkat Inframerah Menggunakan Speech Recognition dengan Spectrogram dan Convolutional Neural Network Berbasis Mikrokontroler," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 5, pp. 955–962, 2023, doi: 10.25126/jtiik.20231056909.

[17] N. Zhang, "Oral Voice Recognition System Based on Deep Neural Network Posteriori Probability Algorithm," *Procedia Comput. Sci.*, vol. 243, pp. 216–223, 2024, doi: 10.1016/j.procs.2024.09.028.

[18] S. Dwijayanti, A. Y. Putri, and B. Y. Suprapto, "Speaker Identification Using a Convolutional Neural Network," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 140–145, 2022, doi: 10.29207/resti.v6i1.3795.

[19] berto nadaek & imam saputra tuminar butar butar, "Implementasi Metode Convolutional Neural Network Untuk Identifikasi Pola Aksara Batak," *Tumiar Butar-Butar | BIMASATI*, vol. 1, no. 2, pp. 49–53, 2022.

[20] N. D. Miranda, L. Novamizanti, and S. Rizal, "Convolutional Neural Network Pada Klasifikasi Sidik Jari Menggunakan Resnet-50," *J. Tek. Inform.*, vol. 1, no. 2, pp. 61–68, 2020, doi: 10.20884/1.jutif.2020.1.2.18.

.