Khazanah Journal of Religion and Technology
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
https://journal.uinsgd.ac.id/kjrt
Online ISSN: 2987-6060
Volume 1 Issue 2 (2023)
DOI: https://doi.org/10.15575/kjrt.v1i2.235

Convolutional Neural Network (CNN) for Detecting Al-Qur'an Reciting and Memorizing

Rizal Hadiyansah
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
1207050109@student.uinsgd.ac.id

Rafi Andamira
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
andamirarafi@gmail.com

Abstract— This research aims to make it easier to memorize the Koran without having to need other people. Memorizers of the Koran (hafiz) often need other people to memorize them to find out if there are errors in their reading. Therefore, this research utilizes machine learning technology to make it easier to read and memorize the Al-Qur'an using the Convolutional Neural Network (CNN) algorithm. CNN was chosen because it is very good at classifying images and audio and can learn and extract features from raw data, such as image and audio data automatically. As a result, the model created succeeded in distinguishing one verse from another very well. The validation results show that the model can correctly detect 57 verses from 64 recorded data, which means it has an accuracy rate of 89.06%. With this verse classification model, it can then be implemented into an application to help memorize the Al-Qur'an even without using the internet.

Keywords- classification, convolutional neural network, memorizing, reciting, Qur'an

I. INTRODUCTION

The Koran is the holy book of all Muslims [1]. Understanding the contents of the Koran is an obligation for all Muslims, including reading and memorizing it [2]. Memorizers of the Qur'an are usually called *hafiz* [3], [4]. *Muroja'ah* is a method of maintaining memorization of the Al-Qur'an by repeating it continuously to achieve mutqin (strength) in reading, memorizing, understanding and practicing [5].

The muroja'ah method is usually done by handing over the memorization to someone else to find out if there are any errors in the memorization. This method is certainly not flexible because of the limitations of other people, especially teachers, who can justify their reading and memorization with a large number of hafiz. Indonesia itself is the country with the largest adherents of Islam in the world with the number reaching 237.56 million people in 2022 [6]. And in 2016, Indonesia was recorded as having around 35 thousand memorizers of the Koran compared to Saudi Arabia which only numbered 6 thousand people [7]. Therefore, hafiz need a new innovation to overcome this problem.

In an increasingly modern era, Muslims are certainly no less innovative in utilizing technology for the benefit of the Ummah. Increasingly developing technology helps make many jobs easier. One thing that is currently at the center of attention is machine learning and Artificial Intelligence. The application of machine learning is not only for data and analysis, but can also be used to process audio. Therefore, this technology can be used to innovate the reading of the Al-Qur'an.

II. RELATED WORKS

Several related studies regarding the reading of the Koran have been carried out using various methods. As in research [7] which solves a similar problem by utilizing a smartphone with the Android operating system. In this research, researchers created an application that utilized the Google Speech API as a voice input medium. The results of testing the application on the surahs Al-Ikhlas, Al-Kautsar, and An-Naas received 100% accuracy. Research has done something similar [8] with test results obtaining an accuracy of 61.1% for appropriate readings and 38.9% for inappropriate readings. Other research [9] also utilized the Google Speech

API to create an Android-based memorization application that produced an accuracy rate of 92%.

Several other studies utilize more renewable and more complex technologies such as machine learning and deep learning. As in research [1] which aims to recognize Al-Qur'an readers by using machine learning with the K-Nearest Neighbor (KNN) classifier algorithm and Artificial Neural Network (ANN) combined with Mel Frequency Cepstral Coefficients (MFCC). The data used amounted to 1930 data for surah Al-Kahf and surah Yasiin. The results of this research were successful in getting 97.03% accuracy for Surah Al-Kahf and 96.08% for Surah Yasiin using KNN, while using ANN, accuracy was 97.6% and 96.7% respectively. Research has done something similar [10] which aims to recognize reciter by using deep learning with the ExtraTreesClassifier, GridSearchCV, RandomizedSearchCV, BLSTM, and SVM with MFCC algorithms producing high accuracy above 98%. On research [11] which aims to differentiate qiraah in reading the Al-Qur'an using Support Vector Machine (SVM) producing an accuracy rate of 96.12% compared to other algorithms and even ANN. This research proves that deep learning is not always better than simpler machine learning algorithms.

Apart from research on reading and memorizing the Al-Qur'an, research on other related topics is also the basis for developing solutions to this problem. As in research [12] which aims to detect the pronunciation of the Arabic alphabet correctly using deep learning with the DCNN, AlexNet, and BLSTM algorithms, producing the highest accuracy reaching 99.14% using AlexNet. Similar research on research [13] with the aim of creating an Automatic Speech Recognition application for the recognition of the East Central dialect of Sundanese using the Deep Neural Network and MFCC methods to produce factors that influence the increase and decrease in the level of accuracy. Research has also done the same thing [14] which aims to create Indonesian speech recognition using MFCC and Recurrent Neural Network (RNN) with 2000 data producing the highest accuracy rate of 73.55%. On research [15] which also aims to create speech recognition using the Linear Predictive Coding (LPC) and K-Nearest Neighbor (K-NN) methods, producing a model accuracy of 62.5%. And on research [16] which aims to find a song by humming the melody of the song using Long Short-Term Memory (LSTM) resulting in an accuracy rate of 50% for the first subject's experiment and 35% for the second subject's experiment.

Several studies that have been reviewed previously are the basis for this research. Researchers are aware of the research [7]–[9] still using the Google Speech API and not yet using an independently created machine learning model specifically for detecting Al-Qur'an reading. Apart from that, research [1], [10] succeeded in creating a model to recognize Al-Qur'an readers but was not designed for general Al-Qur'an reading. And on research [11] which still uses the SVM algorithm, which is not designed for very large amounts of data, such as when all verses in the Al-Qur'an are used as a model. Therefore, researchers in this paper focus on creating a model to detect the reading of the Al-Qur'an by classifying it which is generally designed using a Convolutional Neural Network (CNN). This model is intended to make it easier for hafiz to memorize the Al-Qur'an and maintain istiqomah to continue memorizing without time and place limitations.

III. Research Methods

A. Research Activities

Figure 1 explains research activities. Research begins with identifying the problem, conducting literature studies, collecting data, data preprocessing, modeling, model evaluation, and conclusion results.

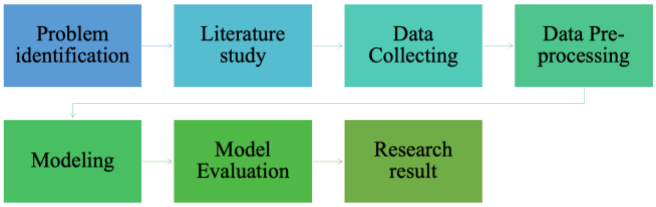

Fig 1. Research Activities

The problem identification process is an effort to find and explain the problem. This process is carried out at the beginning of the research to formulate the problem and find a solution to the problem. Next, a literature study was carried out by collecting journals related to current research.

The data used in this research is data collected through web pages created independently. The web page can be accessed via https://murojaah.azlir.me. The data collected is data from the first surah in the Koran, namely Al-Fatihah, targeting all Muslims of all ages and genders. Researchers are aware of the privacy and security of user data. Therefore, the data collected is general data such as letter number, verse number, voice recording, and delivery time. Apart from that, when storing this data, security has been coated on the server side so that it cannot be accessed by just anyone.

The data that has been collected is then subjected to data preprocessing so that the data is suitable for use. In this process, data is selected and sorted, which will be used in creating the model; data cleaning, by cleaning data from errors, defects and missing or invalid values; and data construction to obtain new data from existing data, which is very important in cases with a minimal amount of data.

At the modeling stage, a model is built using cleaned data. The model that has been trained and validated from the data is then evaluated. At the evaluation stage, the model's accuracy is measured so that differences in the training and testing processes can be seen.

The final stage is model evaluation. At this stage, an evaluation of the training carried out is carried out. This stage evaluates the model using a confusion matrix.

B. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a Neural Network that is often used for image and audio classification

[17]. CNN is designed to be able to learn and extract features from raw data, such as image and audio data automatically.

Audio classification using CNN is different from image classification. In image classification, raw image data with the same dimensions can be used as direct input to CNN. However, in audio classification, time-domain signals which are audio signals can have different dimensions. Additionally, temporal signals often have to be transformed to the frequency domain to reveal unique spectral characteristics, therefore requiring signal transformation [18].

IV. RESULT AND DISSCUSSION

A. Data Collecting

Data collection was carried out through a website created independently. The web page can be accessed via https://murojaah.azlir.me. The appearance of the web page can be seen in Figure 2. Then, the examples of the results of this data collection can be seen in Table 1.

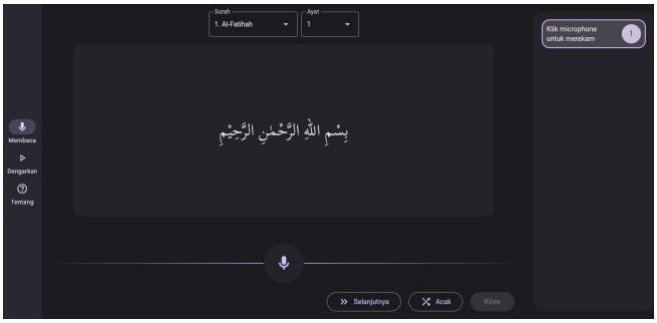

Fig 2. Murojaah web page for data collection and validation

Table 1. Dataset Sample

Surah Number	Ayah Number	URL to Recording	Age	Country	Gender	Qiraah	Recitation Mode	Timestamp of Recording Submission	Has This Recording Been Evaluated?	Label
1	4	https://…	-1	NaN	male	NaN	NaN	2023-06-15 07:11:47.719000+00:00	True	True
1	6	https://…	-1	NaN	male	NaN	NaN	2023-05-28 11:29:18.196000+00:00	True	True
1	5	https://…	-1	NaN	NaN	NaN	NaN	2023-04-08 06:37:41.062000+00:00	True	False
1	5	https://…	-1	NaN	male	NaN	NaN	2023-04-09 21:41:32.592000+00:00	True	True
1	7	https://…	-1	NaN	female	NaN	NaN	2023-04-09 21:41:38.439000+00:00	True	True

B. Data Preprocessing

Data preprocessing is the process of converting raw data into a form that is easy to understand [19], [20]. In this process, the results of data collection are carried out in several stages. The first stage is data validation by labeling the data. Data labeling is done manually by listening to the recording and then labeling it with the appropriate data. Data labeling is done on the same web page as data collection using an admin account. Next, data selection is carried out by selecting the data that will be used and discarding unused data.

The recorded audio data is also processed. Audio data is compressed and the standard sampling rate is adjusted to increase computing efficiency. Apart from that, audio data is augmented, namely creating new data from existing data. This aims to increase the amount of data by making slight differences to the original data. In this research, augmentation was carried out to increase data with four transformations, namely noised, pitch shifted, tanh distorted, and time stretched.

C. Modeling

Model development was carried out using a Sequential model, with a Convolutional Neural Network (CNN) layer. Because the dataset used is not too large, the layers used are also kept simple. This aims to avoid overfitting the model, because if the model is too complex, it will learn the given dataset too precisely, and fail to generalize to data it has not seen before. The layers used in this research can be seen in Figure 3.

The layers used consist of Input Layer, Normalization, Conv2D, Max Pooling 2D, Flatten, Dense, Dense. The Max Pooling layer halves the size of the feature map by shrinking the samples to the maximum value within the window. This is done to avoid a very large number of parameters which can cause the computer to become unstable, and in the end, the model will be too specific to the data. This layer is the reason why CNN can manage large amounts of data in images.

The Dense unit contains hidden layers that relate to the degrees of freedom the model has to try and fit the data. The more complex the data, the more degrees of freedom the model requires. The Flatten Layer combines all the feature map information into one column to be entered into the Dense layer, and finally displays 7 outputs which will be the result of the classification.

Activation functions function to give the model the ability to add non-linearity to the model. The activation function used in this research is the ReLU function. The ReLU function in short means removing negative weights. Meanwhile, the last type of Dense layer activation function uses Softmax, which displays the probability for each class.

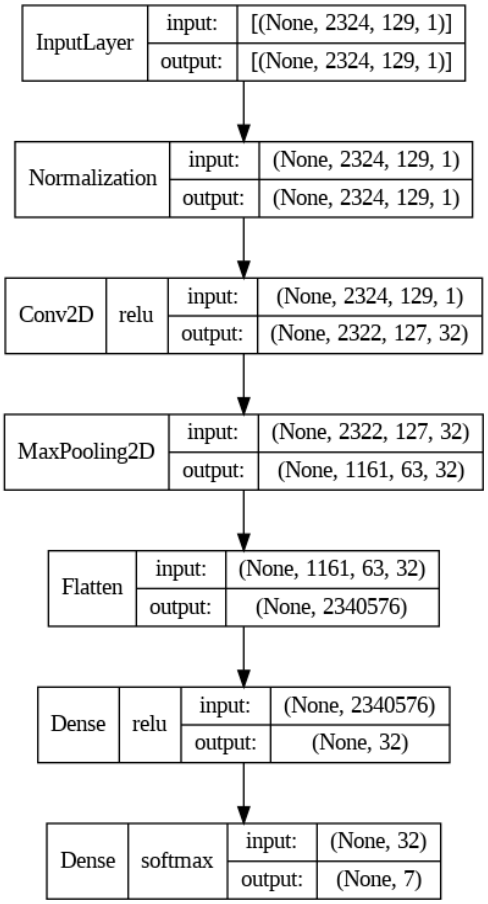

Fig 3. CNN modeling layer

```
model.compile(
    optimizer=tf.keras.optimizers.Adam(),
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=False),
    metrics=['accuracy'],
)
```

Fig 4. Model compilation code using Adam optimizer

The optimizer used is Adam. Adam optimizer functions to manage the learning rate which will evaluate how different the predicted data is from the actual data and will provide penalties for bad prediction results. In this research, researchers used SparseCategoricalCrossentropy for the loss function, where this function is used when each sample belongs to one label, not more than one, and is not a binary classification. This function is suitable for use in this research because each audio sample is included in one category and has a total of 7, namely 7 verses in Surah Al-Fatihah. The evaluation results of this model can be seen in Figure 5.

The graph in Figure 5 shows that the model can learn from the provided dataset. The model is quite good and does not show the characteristics of overfitting and underfitting. The graphs for loss and accuracy both look good with the graph decreasing for loss and increasing for accuracy.

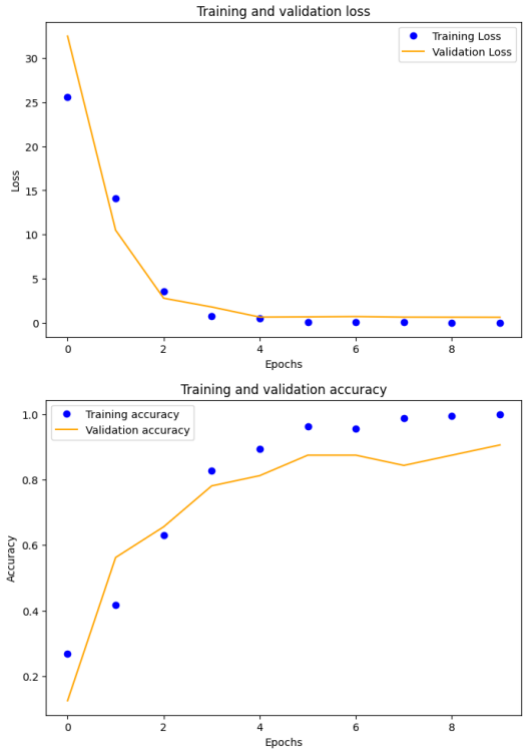

Fig 5. Training and validation result

D. Model Evlauation

The model is evaluated by validating the training results. The model was validated by predicting 64 never-before-seen data sets. The prediction results in the data validation process can be seen using the confusion matrix in Figure 6.

In Figure 6 it can be seen that almost all predictions are correct according to the label. The model succeeded in correctly predicting 57 verses from 64 recorded data. The level of accuracy obtained in this validation process reached 89.06% with a loss of 1.721. This proves that the model successfully learned from the given dataset.

The results obtained in this research were very good with a model accuracy level of 89.06%. However, the level of loss obtained can still be increased with the current loss value at 1.721.

This model was built with a small dataset, namely only for Surah Al-Fatihah. Meanwhile, the Al-Qur'an consists of 114 surahs with more than 6000 verses. Therefore, researchers doubt the reliability of this model when used for very large datasets. This is because the model was built using a classification method per paragraph, not detection per letter.

Apart from that, there are several labels that are not used, such as qiraah, age, gender, and so on due to the lack of dataset and will add to the complexity of the model. Researchers also doubt the reliability of using only audio classification.

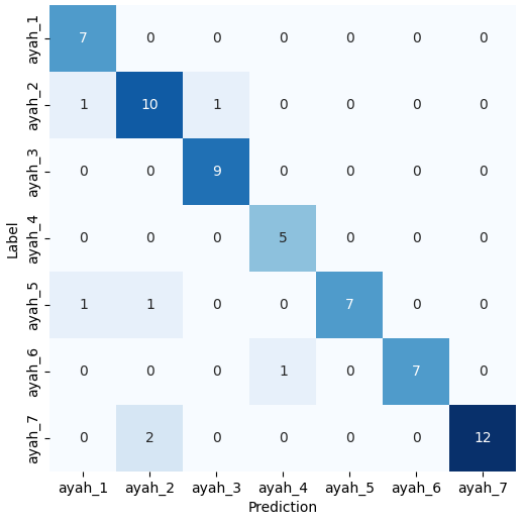

Fig 6. Confusion matrix result

V. CONCLUSION

In this research it has been proven that Convolutional Neural Network (CNN) can be used to classify audio for Al-Qur'an verses. Even though the dataset used is not very large and mixes female and male voices, the model built can still differentiate one verse from another verse well. However, this research is limited to only a small dataset. Therefore, further research is needed for a larger dataset and classification of more verses. Apart from that, this model also cannot correct Tajwid and Makhorijul letters. This is because the method used is a classification so it is not possible to check the tajwid and makhorijul letters. Detection is required per letter by extracting each feature in the Al-Qur'an.

This classification method also does not allow checking verses that are read partially or incompletely. Or in reading by combining more than one verse. These limitations open up further research in the future. Currently, this model can be applied to applications for memorizing each verse of the Al-Qur'an. Because the dataset used is not large, this model can be used locally directly in the application without requiring a server. This is very useful for reducing latency and can be used offline without requiring an internet connection.

REFERENCES

[1] J. H. Alkhateeb, "A Machine Learning Approach for Recognizing the Holy Quran Reciter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, 2020, doi: 10.14569/IJACSA.2020.0110735.

[2] Y. A. Gerhana *dkk.*, "Computer speech recognition to text for recite Holy Quran," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, hlm. 012044, Des 2018, doi: 10.1088/1757-899X/434/1/012044.

[3] "Arti kata hafiz - Kamus Besar Bahasa Indonesia (KBBI) Online." https://kbbi.web.id/hafiz (diakses 27 Mei 2023).

[4] "Definition of HAFIZ." https://www.merriam-webster.com/dictionary/hafiz (diakses 27 Mei 2023).

[5] "Memperbanyak Murojaah | LPK AIKA." https://lpkaika.umt.ac.id/memperbanyak-murojaah/ (diakses 27 Mei 2023).

[6] D. Indonesia, "Jumlah Penduduk Muslim Indonesia Terbesar di Dunia pada 2022," *Dataindonesia.id*. https://dataindonesia.id/ragam/detail/populasi-muslim-indonesia-terbesar-di-dunia-pada-2022 (diakses 27 Mei 2023).

[7] A. Akbar, A. Y. Husodo, A. Zubaidi, dan J. Majapahit, "IMPLEMENTASI GOOGLE SPEECH API PADA APLIKASI KOREKSI HAFALAN AL-QUR'AN BERBASIS ANDROID," vol. 1, no. 1, 2019.

[8] M. Abdurrodjak, M. H. Mud'is, H. Qodim, I. F. S. R. Khaerani, U. Rosidin, dan B. Busro, "Sound matching on the translation of Al-Quran ayat as a learning media for children using mobile-based fast fourier transform and divide conquer algorithm," *J. Phys. Conf. Ser.*, vol. 1402, no. 7, hlm. 077060, Des 2019, doi: 10.1088/1742-6596/1402/7/077060.

[9] M. Assisi, A. Septiarini, A. H. Kridalaksana, dan M. Wati, "Rancang Bangun Aplikasi Hafalan Al-Quran dengan Google Speech API Berbasis Android," *J. Rekayasa Teknol. Inf. JURTI*, vol. 6, no. 1, hlm. 26, Jul 2022, doi: 10.30872/jurti.v6i1.8006.

[10] G. Samara, E. Al-Daoud, N. Swerki, dan D. Alzu'bi, "The Recognition of Holy Qur'an Reciters Using the MFCCs' Technique and Deep Learning," *Adv. Multimed.*, vol. 2023, hlm. 1–14, Mar 2023, doi: 10.1155/2023/2642558.

[11] K. Nahar, R. Khatib, M. Shannaq, dan M. Barhoush, "AN EFFICIENT HOLY QURAN RECITATION RECOGNIZER BASED ON SVM LEARNING MODEL," *Jordanian J. Comput. Inf. Technol.*, no. 0, hlm. 1, 2020, doi: 10.5455/jjcit.71-1593380662.

[12] N. Ziafat, H. F. Ahmad, I. Fatima, M. Zia, A. Alhumam, dan K. Rajpoot, "Correct Pronunciation Detection of the Arabic Alphabet Using Deep Learning," *Appl. Sci.*, vol. 11, no. 6, hlm. 2508, Mar 2021, doi: 10.3390/app11062508.

[13] D. N. Fathurrahman, A. B. Osmond, dan R. E. Saputra, "DEEP NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADA BAHASA SUNDA DIALEK TENGAH TIMUR (MAJALENGKA)".

[14] P. Tridarma dan S. N. Endah, "Pengenalan Ucapan Bahasa Indonesia Menggunakan MFCC dan Recurrent Neural Network," *J. Masy. Inform.*, vol. 11, no. 2, hlm. 36–44, Nov 2020, doi: 10.14710/jmasif.11.2.34874.

[15] M. Walid dan A. K. Darmawan, "Pengenalan Ucapan Menggunakan Metode Linear Predictive Coding (LPC) Dan K-Nearest Neighbor (K-NN)," vol. 7, no. 1, 2017.

[16] H. Hartono, V. C. Mawardi, dan J. Hendryli, "PERANCANGAN SISTEM PENCARIAN LAGU INDONESIA MENGGUNAKAN QUERY BY HUMMING BERBASIS LONG SHORT-TERM MEMORY," *J. Ilmu Komput. Dan Sist. Inf.*, vol. 9, hlm. 106, Jan 2021, doi: 10.24912/jiksi.v9i1.11567.

[17] J. J. Huang dan J. J. A. Leanos, "AclNet: efficient end-to-end audio classification CNN." arXiv, 15 November 2018. Diakses: 22 Juni 2023. [Daring]. Tersedia pada: http://arxiv.org/abs/1811.06669

[18] R. V. Sharan, H. Xiong, dan S. Berkovsky, "Benchmarking Audio Signal Representation Techniques for Classification with Convolutional Neural Networks," *Sensors*, vol. 21, no. 10, hlm. 3434, Mei 2021, doi: 10.3390/s21103434.

[19] A. Wardana dan D. Aldamawan, "Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression," vol. 1, no. 1, 2023.

[20] F. F. Rochdiana dan M. R. I. Darmawan, "The Influence of the Use of Social Media on the Intensity of Worshipping the Millenial Generation using Linear Regression," vol. 1, no. 1, 2023.