JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 8 Number 2 | December 2023: 194-203
DOI: 10.15575/join.v8i2.1081

XGBoost and Convolutional Neural Network Classification Models on Pronunciation of *Hijaiyah* Letters According to *Sanad*

Aaz Muhammad Hafidz Azis[1], Dessi Puji Lestari[2]
[1]School of Computing, Telkom University, Indonesia
[2]School of Electrical Engineering and Informatics, Bandung Institute of Technology, Indonesia

Article Info

Article history:

Received June 08, 2023
Revised October 30, 2023
Accepted October 30, 2023
Published December 28, 2023

Keywords:

CNN,
Hijaiyah,
Pronunciation,
Qur'an.
Sanad,
XGBoost

ABSTRACT

According to *Sanad* (reading designation), the pronunciation of *Hijaiyah* (basics of word and sentence formation in Arabic) letters can serve as a benchmark for correct or valid reading based on the *makhraj* (the place where the letters come out when they are pronounced) and phonetic properties of the letters. However, a scarcity of qualified Qur'anic *Sanad* teachers poses a significant challenge to effective Qur'an learning. This study aims to identify the most practical combination of classification models in constructing a voice recognition system that facilitates Qur'an learning direct teacher interaction. The method used in this research include the XGBoost algorithm and CNN. This research found that the CNN model was employed for 10 out of 12 phonetic property labels, with XGBoost model applied to the remaining two. Furthermore, the inclusion of additional data yielded performance results for each property, with an average accuracy of 78.14% for property S (letters with opposing properties), 70.69% for property T (letters without opposing properties), and an overall average of 73.79% per letter.

Corresponding Author:
Aaz Muhammad Hafidz Azis,
Department of Software Engineering, School of Computing, Telkom University
Jl. Telekomunikasi. 1, Terusan Buahbatu - Bojongsoang, Telkom University, Sukapura, Dayeuhkolot, Kab. Bandung, Jawa Barat 40257
Email: aazmuhammad@telkomuniversity.ac.id

1. INTRODUCTION

There are various types of human life issues mentioned in the Quran. Therefore, our Muslims are responsible for reading, understanding, and implementing its teachings [1]. When reciting the Qur'an, a Muslim should adhere to the recitation and pronunciation taught by the Prophet [2]. Readings acknowledged as authentic or valid and transmitted to Prophet Muhammad are known as readings or *qiraat sanad*. In his work *Jazariyah*, Al-Imam Ibnu Jazariy outlines the conditions for reciting the Qur'an in verse, which include the accurate pronunciation of *hijaiyah* letters and adherence to the *makharijul harfi* [3], [4]. *Hijaiyah* letters constitute the letters found in the composition of the Al-Qur'an. The nature of these letters lies in the distinct qualities that emerge from their *makhraj*. In *Sanad*, the *hijaiyah* letters can serve as a reference for accurate or valid reading, as they embody the essential characteristics of these letters [3].

One significant obstacle to proper learning of the Al-Qur'an is the limited number of qualified instructors. This shortage is particularly evident in the scarcity of Qur'anic instructors in *Sanad*, which hinders the teaching of *tahsin* (the rule of reciting the Al-Qur'an) despite its established standards for pronouncing letters to their characteristics. However, technological advancements have introduced speech recognition systems capable of identifying voices, which are expected to facilitate learning without the need for direct teacher interaction.

Speech recognition is a technology that converts speech signals into digital information. The resulting signal can be processed to recognize sounds at the levels of individual letters, words, and sentences [5]. Several studies have focused on classifying the pronunciation of *hijaiyah* letters based on their *makhraj* and standard Arabic. One such study attempted to develop a model that detects *makhraj* on *hijaiyah* letters using a shallow learning algorithm, specifically Support Vector Machines (SVM) [6]. However, the performance of SVM was found to be lower compared to that of Extreme Gradient Boosting (XGBoost) in sound classification [7], [8]. Additionally, research has explored the comparison of feature extraction between RASTA-PLP in Arabic letter recognition [9]. Durairaj conducted research by combining various extraction techniques, including LPC, MFCC, and RASTA-PLP [10], whereas Helali [11] utilized MFCC, PLP, and LPC extraction techniques.

In 2021, research was conducted on speech recognition systems to classify the pronunciation of *hijaiyah* letters based on their properties [12]. This research utilized the K-Nearest Neighbor (KNN) shallow learning algorithm, resulting in an accuracy value of 66%. Challenges in accurately classifying certain character traits contribute to the relatively low accuracy, subsequently impacting the classification results of *hijaiyah* letters based on their characteristics. The sensitivity of KNN to outlier data influences this limitation.

The characteristics of the KNN algorithm calculates the distance between the feature vectors to be recognized and all stored feature vectors, determining the class most represented among the nearest K feature vectors and storing all the feature vectors, allowing outlier feature vectors to be categorized into other classes [13]. Classifying *hijaiyah* letters based on their properties is challenging, as letter characteristics exhibit more varied and complex pronunciation patterns than classifications based on letter *makhraj* or corpus sounds. Therefore, this research requires feature extraction and classification algorithm advancements to enable more accurate generalizations.

Other studies have demonstrated that using convolutional neural networks (CNN) offers the advantage of generating more accurate generalizations when compared to shallow learning in speech recognition. [14]. *Deep learning* is an algorithm modeled on the structure of the human brain and is composed of numerous layers [15], [16]. Additionally, shallow learning, particularly XGBoost, has shown promise in delivering commendable performance in speech recognition tasks. It has been demonstrated that XGBoost outperforms other shallow learning methods such as SVM, naive Bayes, random forest, and KNN [14], [17]–[19]. Hence, both the XGBoost shallow learning algorithm and the CNN deep learning algorithm have the potential to yield superior results in the classification of *hijaiyah* letters based on their properties.

Based on the research mentioned above [6], [9]-[11], various classification models have been introduced. However, these studies primarily focus on the *makhraj*, or sound of the Arabic corpus, rather than the properties of the letters. Only research [12] utilized KNN for letter trait classification. Nevertheless, it is essential to note that KNN still needs to improve model generalization. Thus, there is a need to develop alternative generalization algorithms, accompanied by a combination of feature extraction, to address these limitations and enhance overall performance.

The primary contribution of this research is to offer the optimal configuration and model for detecting the properties of *hijaiyah* letters. Furthermore, this study has the potential to pave the way for further research in developing and enhancing these classification models. Consequently, these models can be employed as practical learning tools to comprehend the nature of letters in the Quran.

2. METHOD

2.1. System Design

In general, the proposed system design is shown in Figure 1. Figure 1 shows that the system design has the following process:

1. The sound data for pronouncing the *hijaiyah* letters (including data from previous research and newly acquired data) is labeled based on their respective readings.
2. Upon labeling, the voice data undergoes preprocessing, including Voice Activity Detection (VAD) and denoising processes, to eliminate noise from the data.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

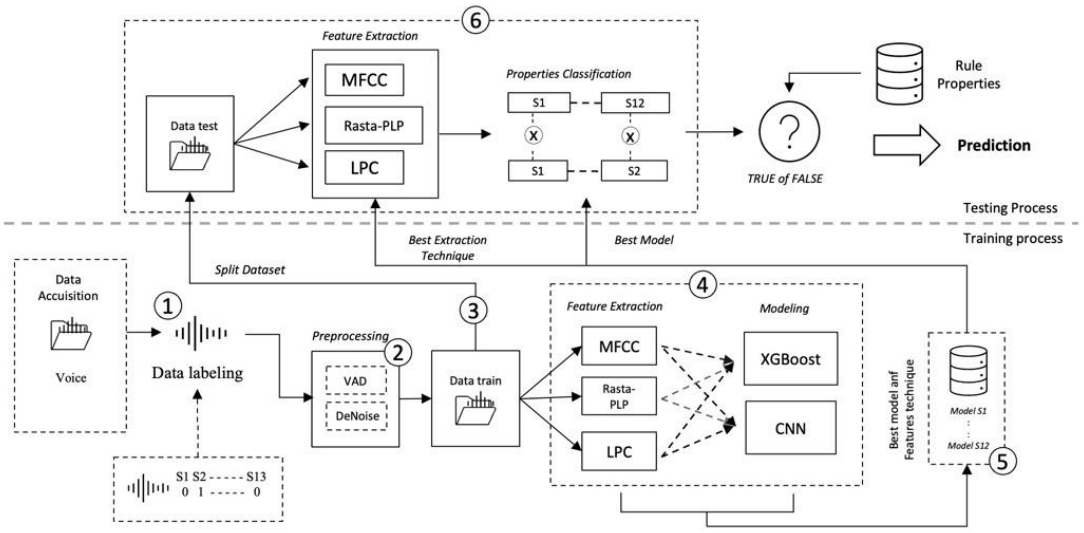

Figure 1. System Design

3. A division is performed between the training data to establish a classification model for the characteristics of *hijaiyah* letters based on their pronunciation and the training data to evaluate the performance of the classification model.
4. During the training stage, the training data undergoes voice data feature extraction. The feature extraction techniques employed include MFCC, Rasta-PLP, and LPC. Subsequently, the feature extraction results serve as input for the XGBoost and CNN classification models. This stage combines feature extraction and classification models to determine the optimal model for characterizing each *hijaiyah* trait.
5. The outcomes of the paired classification model and the best feature extraction configuration for each *hijaiyah* letter characteristic are stored in the storage media.
6. For the testing stage, the test data undergoes voice data feature extraction. Subsequently, the voice data feature extraction results are employed as input for the *hijaiyah* character character classification model. During this phase, the feature extraction model and configuration are retrieved from the storage media, representing the best model and feature extraction from the training process results.

2.2. Data Gathering

Data were gathered from participants who pronounced *hijaiyah* letters based on *Sanad*. This data collection step was implemented to address data scarcity and to manage imbalanced data resulting from incorrect entries. The process of data collection was facilitated using a microphone. The data for this research constituted recordings of participants reading *hijaiyah* letters, all of whom were either following the Sanad or were certified to teach based on the Sanad.

2.3. Preprocessing Data, Feature Extraction, and Model Choosing

A Voice Activity Detection (VAD) process was conducted at the data preprocessing stage to extract pertinent sound data for further processing. Subsequently, the denoising process employed the wavelet denoising method. During the feature extraction phase, sound vectorization was achieved using three distinct techniques, namely MFCC, Rasta-PLP, and LPC. The outcomes of the feature extraction were then normalized before being utilized as input for the classification model. Each feature extraction result was integrated into its respective classification model to identify the model with the best performance. The training process employed a 4-fold cross-validation scheme facilitated by the grid-search technique.

The model architecture utilized in this study was derived from previous research. The CNN classification model's architecture was entirely adapted from the study conducted by Singh and Sharma [14], as illustrated in Figure 2. Subsequently, the model architecture underwent hyperparameter tuning within experimental scenarios designed for each letter characteristic of the *hijaiyah* letters.

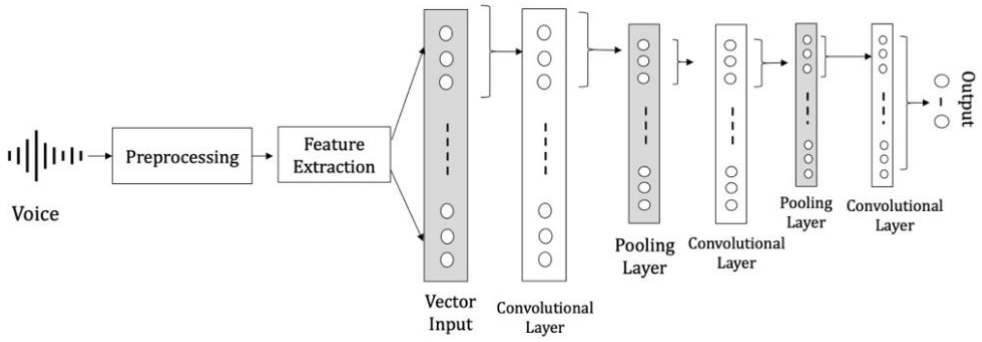

Figure 2 CNN Architecture for sound letters [14]

The CNN architecture incorporateed two layers of the convolutional network, each including a max pooling layer. Furthermore, normalization and dropout batch layers were applied to all convolutional layers. The proposed activation function utilized ReLu for all layers, with the softmax activation function employed in the final layer.

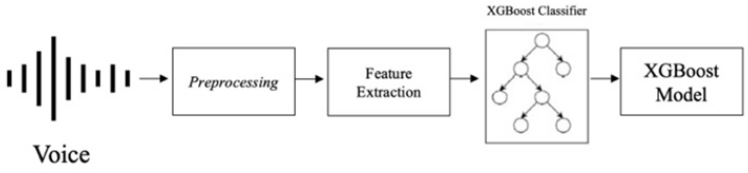

Figure 3. XGBoost Classification Process [18]

In XGBoost, hyperparameter tuning was conducted to identify the optimal hyperplane for its classification. The process flow is depicted in Figure 3. The parameter values were adjusted based on the input feature vectors derived from the sound data of *hijaiyah* letters.

2.4. Data Evaluation and Analysis

The evaluation metric utilized in this thesis research is accuracy. This metric is particularly effective based on the classification distribution between accurate and inaccurate letter pronunciations. The accuracy metric will compare the total number of correct predictions and the total number of predicted data. Precision, recall, and f1-score evaluation metrics are employed to identify misinterpretations and draw conclusions about the model's performance. These evaluations will aid in selecting the best model for classifying the characteristics of each *hijaiyah* letter. Subsequently, an analysis based on these metrics was conducted to determine the optimal model for character classification.

3. RESULT AND DISCUSSION

3.1. Data

The data were gathered from three Qur'an *Sanad* Islamic boarding schools and Quran Tilawatil institutions, namely the Kudang Garut Islamic Boarding School, the Al-Falah Garut Al-Quran Islamic Boarding School, the Al-Qur'an Islamic Boarding School, and the UIN Bandung Tilawatil Qur'an Development Unit. A total of 40 respondents participated in the study, yielding in 21,600 audio recordings captured using a microphone. The recorded audio data had a sampling frequency of 44,100 Hz and duration ranging from approximately 0.5 to 2 seconds. Among the 40 respondents, 20 pronounced the letters correctly, whereas the other 20 mispronounced them.

Each respondent recited 28 *hijaiyah* letters, employig different pronounciations for eacjh letter and repeating each prononciation five times. The variations in reading letters consist of letters with the vowels *fathah*, *kasrah*, and *dhamma*. The process of reading the *hijaiyah* letters is carried out in a *sukun* (ْ) for each *hijaiyah* letter using the help of the hamza letter at the beginning to simulate the sound of the *hijaiyah* letters when *sukun* (ْ).

p-ISSN: 2528-1682
e-ISSN: 2527-9165

3.2. Labelling Data

At this stage, the labeling of the properties for accurate data is presented in Table 1. The labeling process for the properties of inaccurate data involves an inverted or reverse procedure of accurate labeling. This process includes labeling all traits that have opposites or counterparts with the same trait name, commencing with the letter S. Properties that do not have counterparts are labeled with their respective properties, beginning with the letter T. After the labeling, the data undergoes preprocessing using Voice Activity Detection (VAD) and denoising.

Table 1. Properties Of Letters and Their Labels [12]

properties	Label	Properties	Label
Jahr	S1 = 0	*Hams*	S1 = 1
Rakhawah	S2 = 0	*Bayniyyah*	S2 = 1
Istifal	S3 = 0	*Syiddah*	S2 = 2
Isti'la	S3 = 1	*Infitah*	S4 = 0
Idzlaq	S5 = 1	*Ithbaq*	S4 = 1
Ishmat	S5 = 0	*Takrir*	T5 = 1
Qolqolah	T2 = 1	*Shafir*	T1 = 1
Inhirah	T4 = 1	*Li-in*	T3 = 1
Tafasysyi	T6 = 1	*Istithaalah*	T7 = 1

3.3. Preprocessing

There are two steps in the preprocessing stage: Voice Activity Detection (VAD) and denoising. The VAD method utilized is the Giannakopoulos method. Wavelet denoising is employed for the denoising process. These processes yield sound output that exclusively contains essential data and is cleaner than the original data, with reduced noise. The results of the sound data before and after the process are displayed in Figure 4.

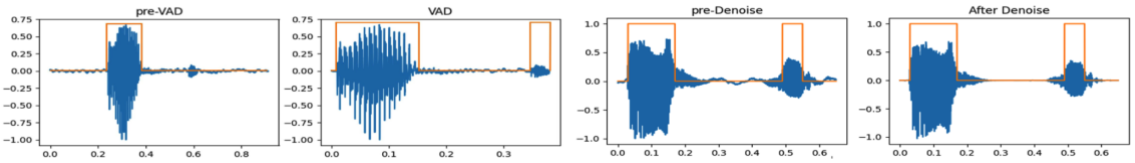

Figure 4. VAD process and Denoising

3.5. Split Test

The dataset distribution ratio is determined based on the respondents. The training data includes recorded voice data from 15 respondents who correctly pronounced the letters and 15 who incorrectly pronounced the letters. Conversely, the test data comprises five respondents who accurately pronounce the letters and five who mispronounce the letters. This distribution ensures an equal percentage of data and prevents the voice data from being mixed with recordings from other respondents. The distribution of data post-division is illustrated in Table 2.

Table 2. Distribution training and data testing

Split Data	Number of sound data
Data training	16.200
Data testing	5.400

3.7. Experiment Scenario

This study conducted three experimental scenarios, as depicted in Table 3. The best model from the experimental results will be determined and summarized to demonstrate the entire system's performance in detecting the accuracy of each *hijaiyah* letter character. The explanation of each experimental scenario is as follows:

1. **Experiment P1**. In this experiment, feature extraction was compared, considering variations in MFCC, LPC, and Rasta-PLP. The experiment will be executed using grid-search in conjunction with the P2 experiment.
2. **Experiment P2**. In this experiment, the pursuit of the optimal classification model was undertaken. The model variations encompass XGBoost and CNN. These two models are coupled with the feature extraction utilized in the P1 experiment, enabling each model to access three

distinct types of feature extraction. Additionally, this experiment aims to determine the most effective hyperparameters for each type of classification model. The validation scheme employed in this experiment involved 4-fold cross-validation using the accuracy metric.

3. **Experiment P3**. This experiment involves taking the classification model with the best accuracy results from a combination of feature extraction, model, and hyperparameters in each model within experiment P2. The model is then retrained to classify each character per letter, employing a multi-model classification technique for each trait. Additionally, this experiment varied the number of epochs and estimators.

Table 3. Experiment Scenario

ID	Variable		Value	Number of Combination
P1	Feature Extraction		MFCC, LPC, Rasta-PLP	3
P2	Hyperparameter model	XGBoost	Minimal Child Weight: {1, 5, 10,15}	3
		CNN	Optimizer :{adam, SGD, RMSProp}	8
			Learning Rate : {0.01, 0,001}	
	Number of XGBoost Experiment			**12**
	Number of CNN Experiment			**18**
P3	XGBoost		N_Estimator:{300, 500}	2
	CNN		Epoch: {100,300,400}	3
	Experiment Total			**35**

The best model, as determined by the experimental results, will evaluate its performance for each *hijaiyah* letter. The performance outcomes will be summarized to demonstrate the system's efficacy in accurately detecting each *hijaiyah* letter.

3.8. Result and Analysis

The analysis process is divided into three stages: feature extraction analysis, classification model analysis, and letter-specific model analysis.

3.2.1. Extraction Feature Analysis

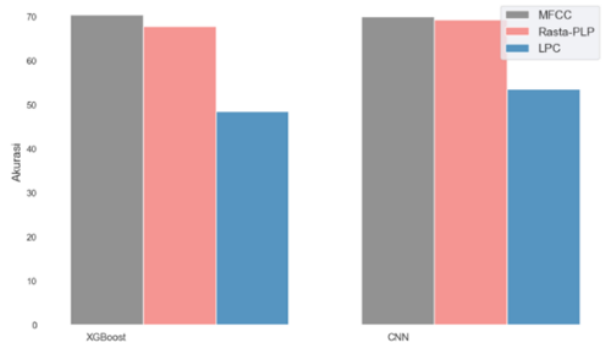

Figure 5. Comparison of Extraction Feature Accuracy

Based on the results of experiments P1 and P2 using the grid search technique, Figure 5 depicts the average accuracy of the fine-tuning process involving feature extraction, classification models, and hyperparameters. The MFCC feature extraction technique demonstrates an advantage in providing representative features compared to the other two techniques. However, Rasta-PLP also yields features that exhibit only marginal differences from MFCC, with slightly superior accuracy in specific aspects. While the MFCC technique outperforms Rasta-PLP in certain characteristics, the disparity between the two methods is relatively minimal.

Both MFCC and Rasta-PLP offer fairly representative features for classifying the pronunciation of *hijaiyah* letters based on their nature. This is attributed to their compatibility with the human hearing system and superior feature extraction capability, especially compared to LPC. Additionally, MFCC displays a lower level of robustness compared to Rasta-PLP. Notably, in this study, a denoising process was initially conducted, ensuring that the data used was sufficiently cleansed of noise, enabling MFCC to outperform Rasta-PLP despite the latter's robustness against noise. On the other hand, LPC exhibits less efficacy in distinguishing between letters that share similarities with other letters.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

3.2.2. Model analysis

Table 4. Best tuning result for each property

	XGBoost			CNN			
Label	**Feature Extraction**	**Min Child Weight**	**Accuracy**	**Feature Extraction**	**LR**	**Optimizer**	**CNN**
S1	MFCC	15	81.28%	MFCC	0.001	Adam	82.20%
S2	MFCC	5	75.98%	RPLP	0.001	Adam	76.93%
S3	MFCC	5	73.19%	MFCC	0.001	Adam	72.44%
S4	RPLP	10	78.58%	MFCC	0.01	Sgd	83.89%
S5	MFCC	10	75.50%	MFCC	0.001	Adam	77.37%
T1	RPLP	5	66.81%	RPLP	0.001	Adam	63.17%
T2	MFCC	10	85.43%	MFCC	0.001	Adam	88.26%
T3	MFCC	1	60.80%	MFCC	0.01	Sgd	62.76%
T4	RPLP	5	61.50%	MFCC	0.01	Sgd	63.85%
T5	MFCC	10	72.59%	RPLP	0.001	Adam	74.93%
C	MFCC	10	69.19%	MFCC	0.01	Sgd	71.91%
T7	MFCC	1	65.98%	MFCC	0.01	Adam	63.48%
Average			**72.23%**				**73.72%**

Table 4 illustrates the accuracy of each value obtained from the best model using the XGBoost and CNN classification algorithms. According to the results presented in Table 4, after hyperparameter tuning, the CNN algorithm model demonstrates superior performance compared to XGBoost. The CNN classification outperforms the XGBoost model in 10 of 12 labels, specifically S1, S2, S4, S5, T1, T2, T3, T4, T5, and T6. However, the XGBoost model exhibits better accuracy for the S3 and T7 labels, suggesting that the XGBoost algorithm encounters challenges in learning voice data. This trend is reflected in the overall average accuracy values, which reach 72.23% for the XGBoost model with the best tuning using an MFCC and Min Child Weight value of 10. Meanwhile, the CNN model achieves an overall average accuracy of 73.72% with the best tuning parameters set to LR 0.001, MFCC, and Adam optimizer.

3.2.3. Letters Classification Analysis

Analysis per letter was conducted using the optimal model derived from experimental results employing multi-model classification modeling. The best model's test results are assessed based on the letter properties. Therefore, the output of this classification indicates the accuracy of the classification, whether the properties are present or absent, and whether the classification is correct or incorrect.

Table 5. Index Score [20]

Index Score	**Accuracy**
Good	81%-100%
Enough	61%-80%
Not Enough	41%-60%
Bad	<40%

Based on the standard assessment reference in Table 5, two groups have been formed according to the achieved accuracy. The outcomes of this grouping are presented in Table 6.

Table 6. Accuracy Average Result based on Properties

Index Score	**Letters**	**Average**
Good	ب (ba'), ج (Jim), ق (Qaf)	73,79%
Enough	ا (alif), ب (ba'), ت (ta'), ج (jim), ح (ha'), د (dal), ط (tha'), غ (ghain), ل (lam), ث (tsa), خ (kha'), ذ (dzal), ر (ra'), ز (za), س (sin'), ش (syin), 14. ص (shad), 17. ظ (zha'), ع ('ain), غ (ghain), ف (fa'), ق (qaf), 25. ن (nun), و (wau), هـ (haa), ي (ya')	

In Table 6, three letters are classified in the 'good' category, with an accuracy exceeding 80%. In the 'sufficient' category, the attained accuracy surpasses 60%. The overall average accuracy achieved is 73.79%. Then, samples were selected based on the letters with the highest accuracy, such as the letter *ba* (ب), and those with the lowest accuracy, such as the letter *dhod* (ض), and subsequently, the classification results were evaluated.

Table 7. Performance Model on *Ba and Dhod*

	Ba (2)				*Dhad (15)*			
	Accuracy	**Precision**	***Recall***	**F1**	**Accuracy**	**Precision**	***Recall***	**F1**
S1	**86.64%**	94.59%	81.40%	87.50%	**52.22%**	84.62%	40.74%	55.00%
S2	**87.78%**	86.48%	86.08%	86.28%	**74.07%**	70.90%	72.83%	71.85%
S3	**83.93%**	79.55%	92.11%	85.37%	**65.13%**	89.66%	53.06%	66.67%
S4	**87.70%**	92.31%	85.71%	88.89%	**71.06%**	96.15%	55.56%	70.42%
S5	**75.06%**	91.18%	65.96%	76.54%	**62.39%**	96.55%	50.91%	66.67%
T1	**74.28%**	97.22%	64.81%	77.78%	**47.16%**	57.89%	46.81%	51.76%
T2	**96.10%**	97.44%	95.00%	96.20%	**82.84%**	84.62%	82.50%	83.54%
T3	**81.44%**	96.15%	65.79%	78.13%	**57.45%**	64.86%	55.81%	60.00%
T4	**82.98%**	80.00%	90.00%	84.71%	**60.78%**	65.85%	62.79%	64.29%
T5	**72.29%**	88.10%	69.81%	77.89%	**47.97%**	58.70%	57.45%	58.06%
T6	**71.83%**	75.00%	69.23%	72.00%	**47.95%**	60.87%	56.00%	58.33%
T7	**82.90%**	80.00%	82.35%	81.16%	**75.96%**	80.95%	77.27%	79.07%

Referring to Table 7, it is evident that the average precision value for the letter ب (*ba'*) is at a satisfactory level, while the recall for the same letter remains sufficient. This implies the existence of data deemed accurate despite being incorrect, resulting in an F1 score of 81.16%.

Table 8. Properties Accuracy Average for each letter

Properties of letters	**Score average**	**Properties of letters**	**Score average**
S1	52.22	T1	47.16
S2	74.07	T2	82.84
S3	65.13	T3	57.45
S4	71.06	T4	60.78
S5	62.39	T5	47.97
		T6	47.95
		T7	75.96

The model for the letter ض (*dhad*) still struggles to effectively classify its properties, as demonstrated by the low accuracy, precision, recall, and f1-score. The model's classification performance remains subpar, especially in terms of recall. This underperformance in recall can be attributed to many erroneous data points being erroneously categorized as correct. This issue aligns with similarities between accurate ض (*dhad*) letters and inaccurate ض (*dhad*) letters.

Regarding letter properties, Table 8 illustrates the average value per letter. Properties with opposing characteristics (S1 to S5) exhibit an accuracy exceeding 70%. Conversely, properties without opposing characteristics (T1-T7) tend to display lower accuracy than their counterparts. This disparity arises from two factors, namely:

a. This training scheme encompasses traits that do not possess opposites, namely the properties of T1-T7. Moreover, the training process still includes letter data, which should be unnecessary in the training process, despite these properties being exclusive to specific letters, as exemplified by the following properties: (1) T1 (*qolqolah*), exclusive to the letters *ba, Jim, dal, tho, qof, kaf.* (2) *T2 (liin),* exclusively applicable to the letters ya and wawu.
b. Data labeling errors are not inherent but rather a result of mislabelling. Labeling was conducted through an inverse process based on authentic data in this study. For instance, the letter 'ba' (2) exhibits the characteristics of *jahr, syiddah, istifal, infithah, idzlaq*, and qolqolah. Labeling involves identifying the sound data for '*ba*' that solely represents *each* characteristic, '*ba*' without the *qolqolah* characteristic, and *'ba'* without the subsequent characteristic.

Characteristic T2 (*qolqolah*) has an accuracy of 77%. This is because *Qolqolah* letters have characteristics that are quite easy to classify. Namely, the characteristics of the data in the trait have a long enough duration after cutting. When the data contained in the data is wrong, the process is carried out by eliminating the reflection in the sound. In the *Normalization* process, the data that has *Qolqolah* properties, there is silence, but the silence is considered as sound so that it can be seen from the duration and characteristics of the sound; it can be distinguished quite well compared to data that does not have *qolqolah*. With the VAD, only voice data is taken, which is important. The duration will also be cut off at the beginning and end. However, when it is silent in the middle, the data will not be cut off, so important data on the sound with *qolqolah* properties tend to be longer in duration than not *qolqolah*.

The S2 trait has lower accuracy compared to the other S traits; this is because this trait has three classes, which are *Syiddah, bayiniyah*, and *rakhawah* traits. These three properties become difficult due to the nature of the *bayiniyah* (middle), which requires the pronunciation model of the letters to capture the character of the sound in great detail and classify these letters into the *bayniyyah* class. With these three classes, the model still has difficulty getting good results because of the high accuracy required to capture the sound character.

4. CONCLUSION

This research has presented the development of a combination of classification models for *hijaiyah* letter traits using CNN and Gradient Boosting (XGBoost) algorithms in conjunction with MFCC, RASTA-PLP, and LPC feature extraction. Tests employing these three feature extraction methods reveal that MFCC excels in the majority of traits. Furthermore, the CNN algorithm demonstrates superior performance compared to the XGBoost algorithm for traits S1, S2, S4, S5, T1, T2, T3, T4, T5, and T6, whereas the XGBoost model excels in labeling traits S3 and T7. Based on the best combination of models, hyper-tuning, and additional data integration, the accuracy for S properties averages 78.14%, while T properties achieve 70.69%. The overall average accuracy per letter stands at 73.79%. In future research, it is recommended to eliminate non-target sound data to avoid bias in the classification model. Future model should be capable of recognizing readings or verses from the Qur'an by the *Sanad* readings.

REFERENCES

[1] K. Robbayani, "The Position And Position Of Al Quran As Islamic Law Sources," *Proceeding International Seminar on Islamic Studies*, vol. 1, 2019.

[2] A. Hakim and O. P. Pratama, "Al-afro As-Sabah and Its Relationship with Al'Qira'at; Theory and Refutation of Orientalist Criticism of the Qur'an," *Takwil: Journal of Quran and Hadith Studies*, vol. 1, no. 1, pp. 17–31, Jun. 2022, doi: 10.32939/two.v1i1.1256.

[3] A. Al-Fadhili, *Tajwidul Quran Metode Al-Jazary*, 1st ed., vol. II. Bandung: Tajwid Online, 2017. Accessed: Feb 26, 2022. [Online]. Available: www.tlgrm.me/online_tajwid

[4] M. Darwis Hude, A. S. Muhammad, and S. Sunarsa, "Penelurusan Kualitas dan Kuantitas Sanad Qiraah Sab'ah: Kajian Takhrij Sanad Qiraah Sab'ah," in *Sasa Sunarsa Misykat*, Misykat, 2020.

[5] H. D. Shah, A. Sundas, and S. Sharma, "Controlling Email System Using Audio with Speech Recognition and Text to Speech," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Sep. 2021, pp. 1–7. doi: 10.1109/ICRITO51393.2021.9596293.

[6] L. Marlina *et al.*, "Makhraj recognition of Hijaiyah letter for children based on Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machines (SVM) method," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 935–940. doi: 10.1109/ICOIACT.2018.8350684.

[7] S. S. Shanta, M. Sham-E-Ansari, A. I. Chowdhury, M. M. Shahriar, and M. K. Hasan, "A Comparative Analysis of Different Approach for Basic Emotions Recognition from Speech," in *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*, IEEE, Sep. 2021, pp. 1–4. doi: 10.1109/ICECIT54077.2021.9641208.

[8] D. A. Rahman and D. P. Lestari, "COVID-19 Classification Using Cough Sounds," in *Proceedings - 2021 8th International Conference on Advanced Informatics: Concepts, Theory, and Application, ICAICTA 2021*, IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/ICAICTA53211.2021.9640278.

[9] J. Farooq and M. Imran, "Mispronunciation Detection in Articulation Points of Arabic Letters using Machine Learning," in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2021, pp. 1–6. doi: 10.1109/ICECube53880.2021.9628251.

[10] P. Durairaj and R. Shyamala, "A Review On Performance Of Voice Feature Extraction Techniques," 2019, pp. 221–231. doi: 10.1109/ICCCT2.2019.8824988.

[11] W. Helali, Z. Hajaiej, and A. Cherif, "Arabic corpus implementation: Application to speech recognition," in *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, IEEE, Mar. 2018, pp. 50–53. doi: 10.1109/ASET.2018.8379833.

[12] M. Ridha, "Identifikasi Kebenaran Bacaan Huruf Hijaiyah Sesuai Sanad Menggunakan Metode MFCC dan K-Nearest Neighbor (KNN)," Universitas Gadjah Mada, Yogyakarta, 2021.

[13] S. Hazmoune, F. Bougamouza, S. Mazouzi, and M. Benmohammed, "A new hybrid framework based on Hidden Markov models and K-nearest neighbors for speech recognition," *Int J Speech Technol*, vol. 21, no. 3, pp. 689–704, 2018, doi: 10.1007/s10772-018-9535-4.

[14] V. Singh and K. Sharma, "Empirical Analysis of Shallow and Deep Architecture Classifiers on Emotion Recognition from Speech," in *Proceedings - 6th IEEE International Conference on Cyber Security and Cloud Computing, CSCloud 2019 and 5th IEEE International Conference on Edge Computing and Scalable Cloud, EdgeCom 2019*, Institute of Electrical and Electronics Engineers Inc., Jun. 2019, pp. 69–73. doi: 10.1109/CSCloud/EdgeCom.2019.00-16.

[15] D. Wang, X. Wang, and S. Lv, "An overview of end-to-end automatic speech recognition," *Symmetry*, vol. 11, no. 8. 2019. doi: 10.3390/sym11081018.

[16] D. Jakhar and I. Kaur, "Artificial intelligence, machine learning, and deep learning: definitions and differences," *Clinical and Experimental Dermatology*, vol. 45, no. 1. 2020. doi: 10.1111/ced.14029.

[17] H. Chen, Z. Liu, X. Kang, S. Nishide, and F. Ren, "Investigating voice features for Speech emotion recognition based on four kinds of machine learning methods," in *2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 2019, pp. 195–199. doi: 10.1109/CCIS48116.2019.9073725.

[18] J. V. Egas-López and G. Gosztolya, "Predicting a Cold from Speech Using Fisher Vectors; SVM and XGBoost as Classifiers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12335 LNAI, 2020, pp. 145–155. doi: 10.1007/978-3-030-60276-5_15.

[19] M. J. Al Dujaili, A. Ebrahimi-Moghadam, and A. Fatlawi, "Speech emotion recognition based on SVM and KNN classifications fusion," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, 2021, doi: 10.11591/ijece.v11i2.pp1259-1264.

[20] A. Zaidah and E. Mahariyanti, "Validation of Scientific Learning," *Path of Science*, vol. 6, no. 12, pp. 3007–3011, Dec. 2020, doi: 10.22178/pos.65-5.