Scope
Volume 14 Number 01 March 2024

Effects of Data Set Size on Speech Classification

O.K. Adejumobi[1*], A.I.O. Yussuff[2] & A. A. Adenowo[2]
[1]The Polytechnic, Ibadan, Nigeria, Department of Computer Engineering
[2]Lagos State University, Nigeria, Department of Electronic and Computer Engineering
*Corresponding author: O. K Adejumobi

Abstract : This paper determined the effects of dataset size on theaccuracy of dialects classification models. To achieve this aim, an experimental methodology, where two (2) datasets A and B of varying sizes were used. Dataset A has a total number of 500 samples (100 samples for each of the classes) while Dataset B has a total number of 7000 samples (1400 samples for each of the classes). Both datasets were divided into; 70%, for network training, 20%, for validation and 10%, for prediction. The datasets contain audio samples of Egba, Ekiti, Ibadan, Ijebu and Ondo dialects collected from participants via mobile phones, radio and sound recorders. A Convolutional Neural Network (CNN) Classifier was developed.The process of achieving the objective of this research was divided into four (4) main stages namely: speech signals acquisition, data pre-processing, speech data classification and Model training/ testing and evaluation. The Model was implemented on Matlab 2022b platform. With the same Classifier, the results showed that the larger sized dataset 'B' gave a better performance accuracy of 100% for all the dialects classes. While the smallerdataset 'A' gave a performance accuracy of the Model's predictions for Egba, Ekiti, Ibadan, Ijebu and Ondo as 98.8%, 98.2%, 96.8%, 95.1% and 97.4% respectively. However, it is recommended that the complexity of the Model be considered before increasing the datasets to avoid under-fitting of the network.

Keywords: Accuracy,Classification, Convolutional Neural Network, Data Sets, Dialects, Iteration, Model, Over-fitting, Speech and Under-fitting.

I. Introduction

Speech classification is a challenging task with small datasets, Rahmanand Sultana (2017) revealed 'that large datasets lead to better classification performances while small datasets trigger over-fitting and unreliable biased classification models." However, in some healthcare services data collection faces many challenges (Mehrafarin et al 2022) due to lack of cases according to Marcoulides (2005) as well as legal challenges Wieczorek (2019). In the medical domain, Alhanoof et al (2021) investigated 'the impact of dataset size on the performance of supervised machine learning models using small and large datasets'. The results showed great improvement with large dataset.

Further studies also investigated the extent to which dataset size (Dris et al 2019), impact the classification performance inobject detection (Zhu et al, 2019), sentiment classification (Choi and Lee, 2017), information retrieval (Linjordet and Balog, 2019) and plant disease classification (Barbedo, 2018).

II. Methodology

A Convolutional Neural Network (CNN) Classifier was developed. The developed Model is divided into four (4) main stages namely: speech signal acquisition, data pre-processing, speech data classification and model evaluation (see Figure 1).

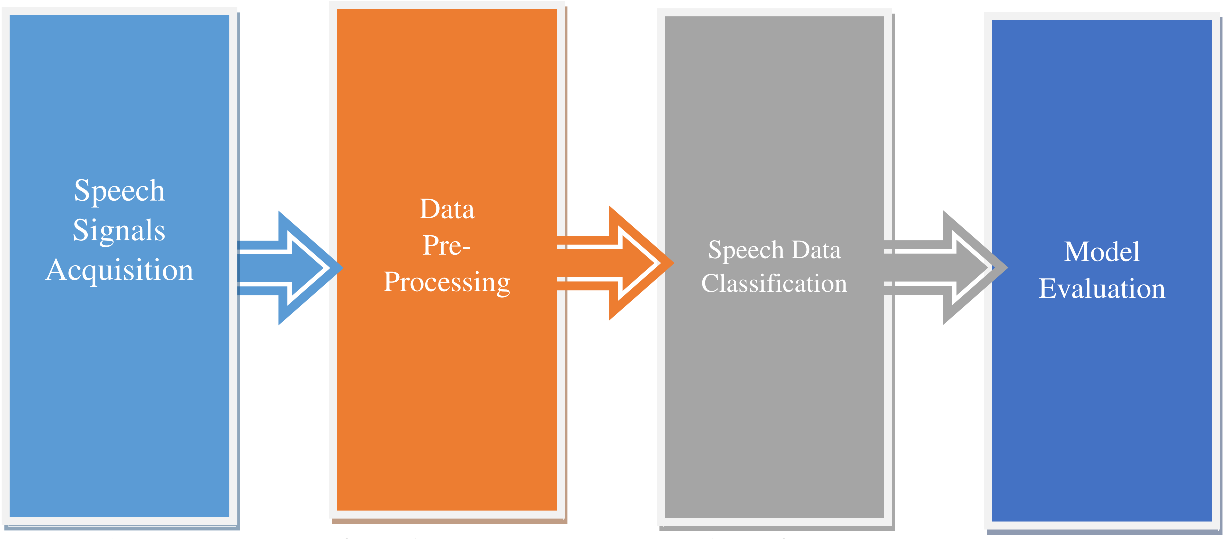


Figure 1: Block Diagram of Dialects Recognition/Classification System

1. *Speech Signal Acquisition*

Samples of five (5) Yoruba dialects were obtained namely; Ibadan, Ijebu, Egba, Ekiti and Ondo. The dataset were recorded at different environments, sample rate and styles. Two (2) data sets, A and B were obtained in "Opus file" format. Dataset A contains 100 samples of each of the five classes making a total 500 dialects samples while B contains 1,400 samples of each dialect making a total of 7000 dialects.

2. *Data Pre-processing*

To prepare the datasets for efficient training using CNN, they were ee first converted to ".wav" format using EZ CD audio Converter Software and to auditory-based spectrograms (Figures 2 and 3).

3. *CNN – Based DialectsClassification Process*

A deep learning network was developed to classify the dialects signals into five (5) classes. The block diagram for the developed dialect classification Model is shown in Figure 4.

4. *Performance Evaluation of the Developed Classification Model.*

Confusion Matrix was used to determine the correctness of the Model. The Model was evaluated using accuracy(equation 1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Where,

TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

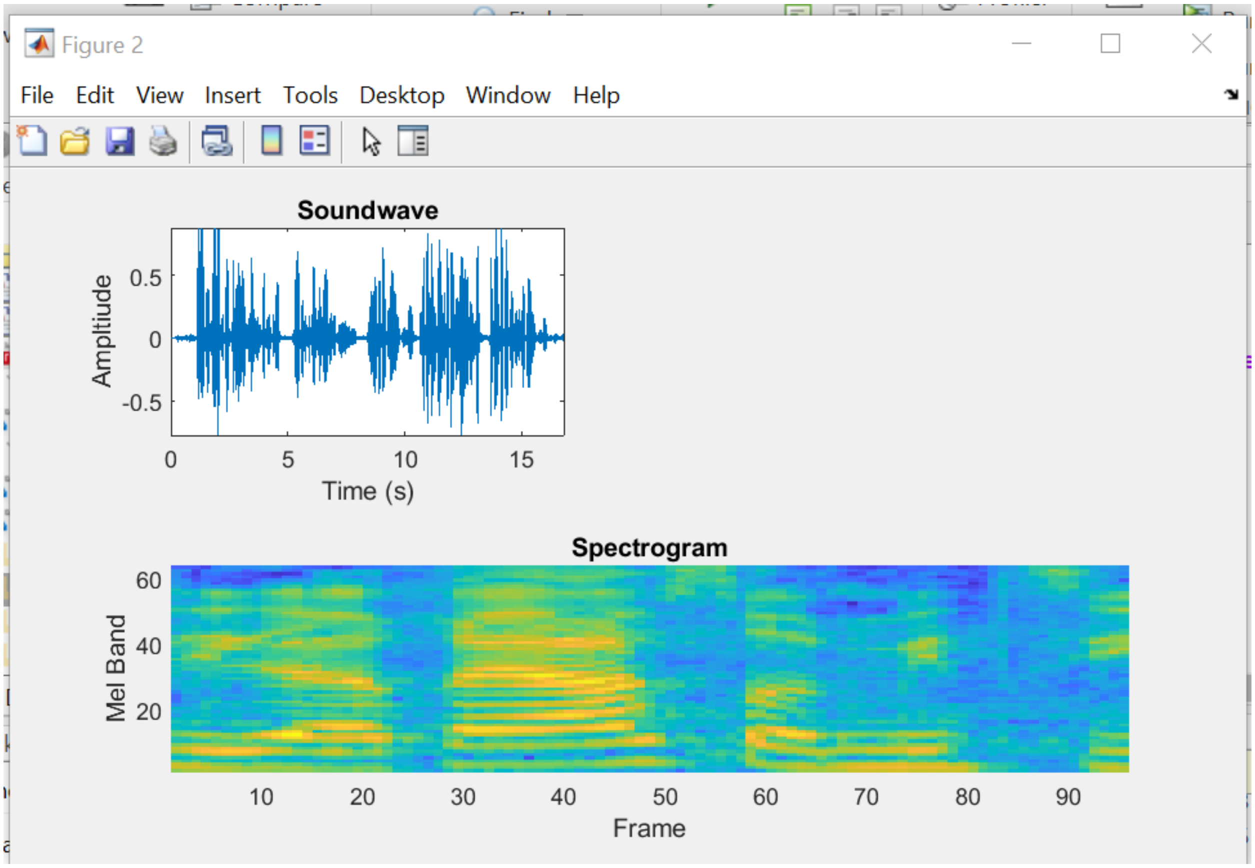


Figure 2: Input Sound wave and Spectrograms for Dataset A.

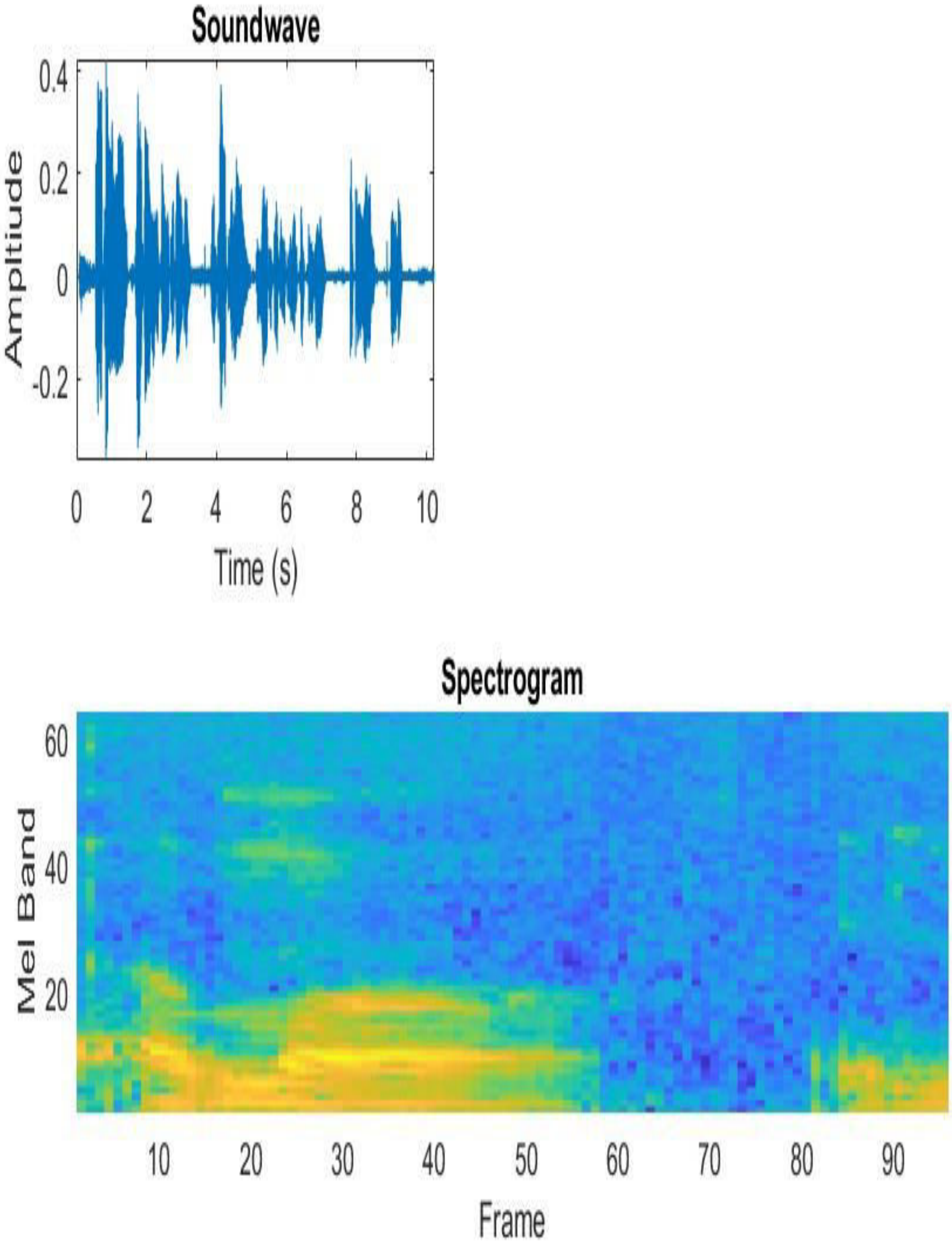


Figure 3: Input Sound wave and Spectrograms for Dataset B.

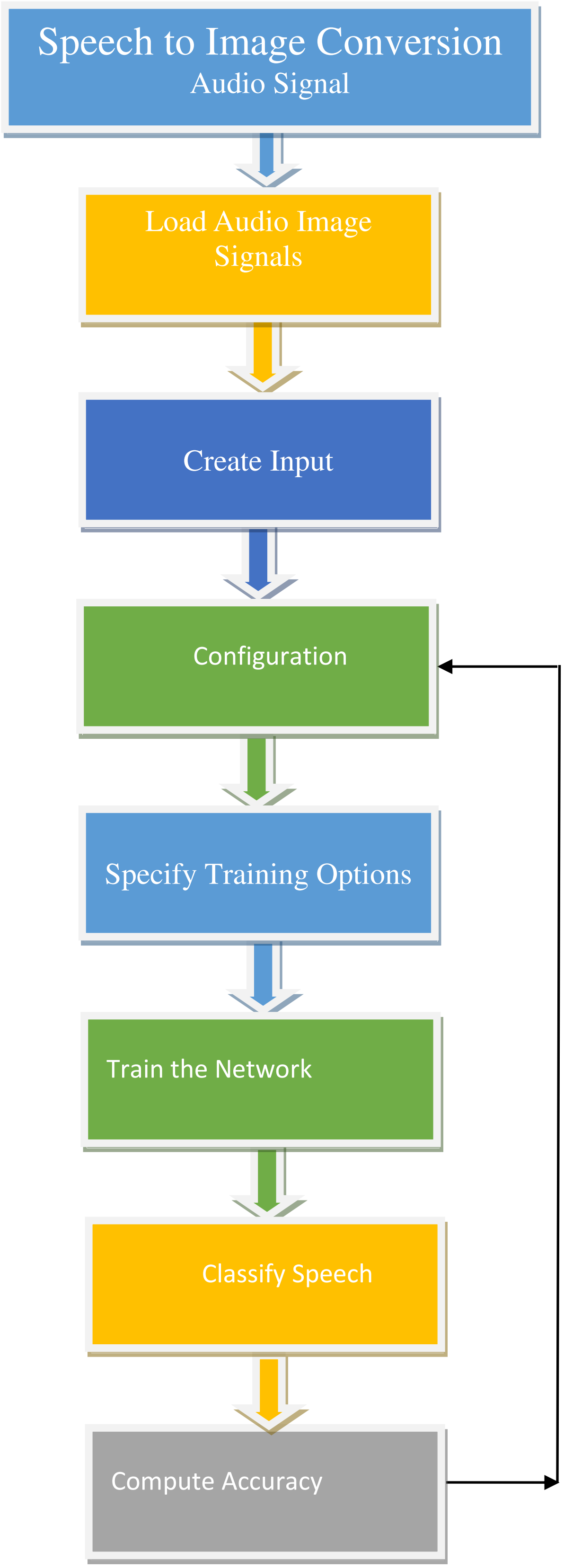

Speech to Image Conversion
Audio Signal
Load Audio Image Signals
Create Input
Configuration
Specify Training Options
Train the Network
Classify Speech
Compute Accuracy

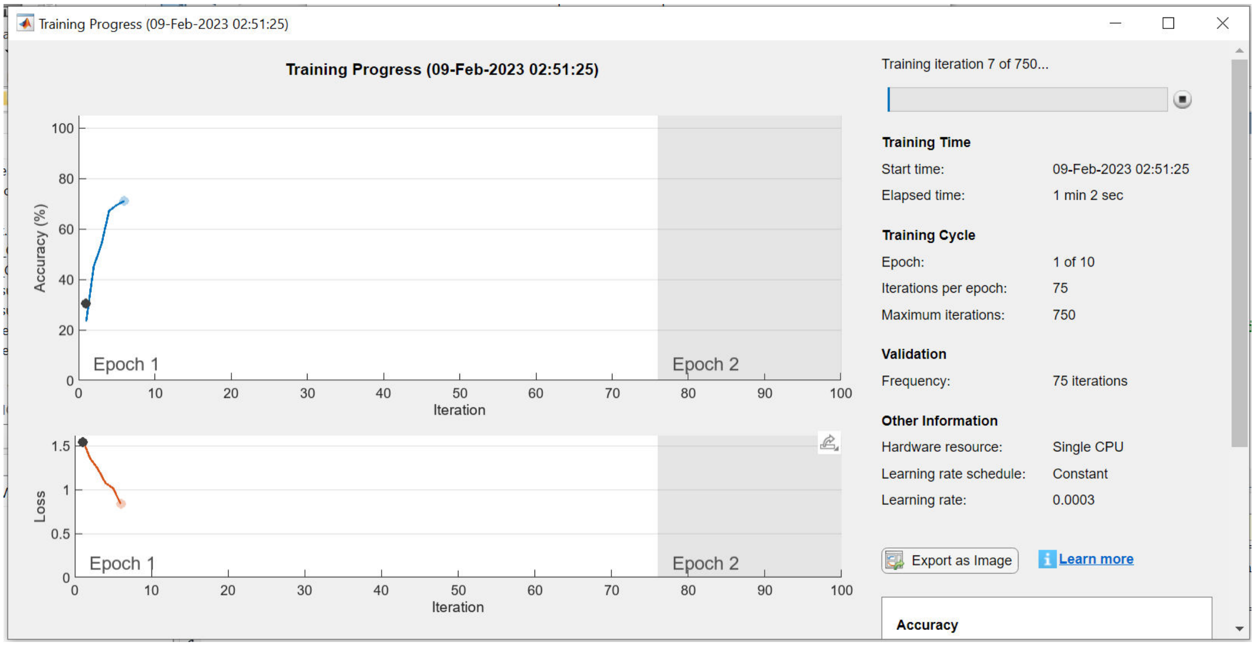
Figure 4: Block Diagram for the Developed Dialect Classification Model.

III. Results

The experimental results are presented for the classification model with both small datasets (A) and larger datasets (B). The experiments were carried out on Matlab 2022b platform.

1.Results of the Network Training Section for Dataset A.

70% of the datasets of each dialect class were used in training the network. The speech signals obtained were classified into five dialect classes using the CNN.Figures 5 to 9 show the progressive training graphs of 750 iterations for the classification of the speech Signals.



Training iteration 7 of 750...

Training Time
Start time: 09-Feb-2023 02:51:25
Elapsed time: 1 min 2 sec

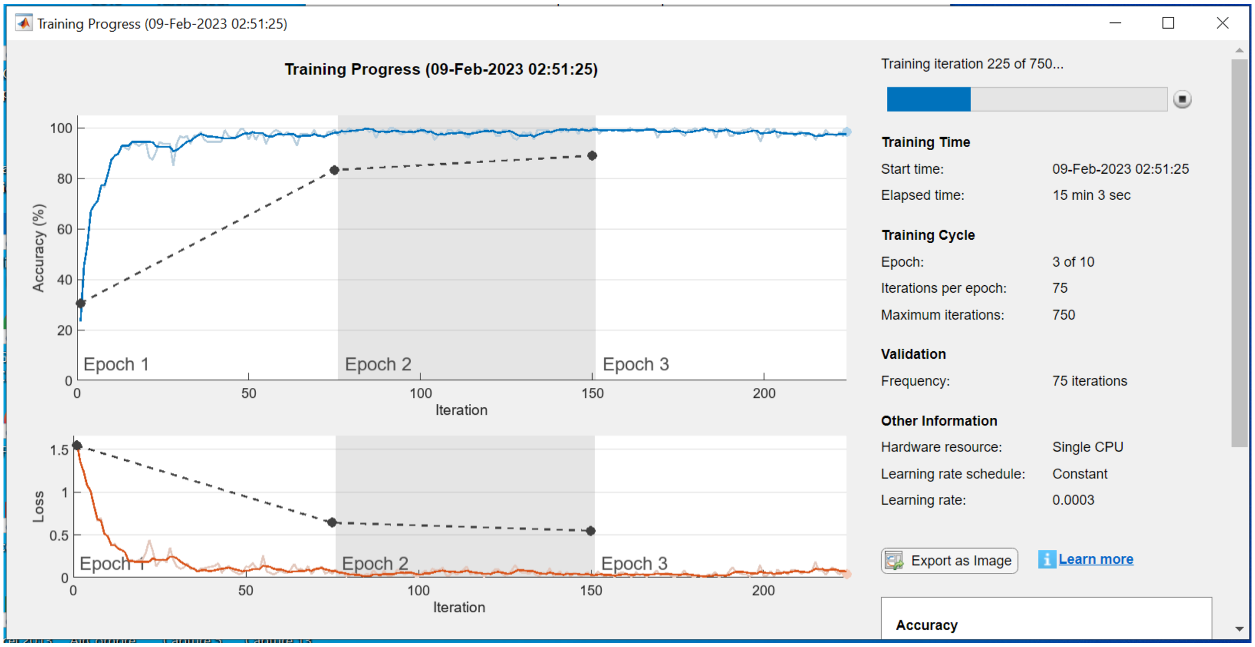
Training Cycle
Epoch: 1 of 10
Iterations per epoch: 75
Maximum iterations: 750

Validation
Frequency: 75 iterations

Other Information
Hardware resource: Single CPU
Learning rate schedule: Constant
Learning rate: 0.0003

Figure 5: Training Progressive Graph for the Developed CNN Model at Iteration 7 of 750.

-


Training Progress (09-Feb-2023 02:51:25)

Training Progress (09-Feb-2023 02:51:25)

Training iteration 225 of 750...

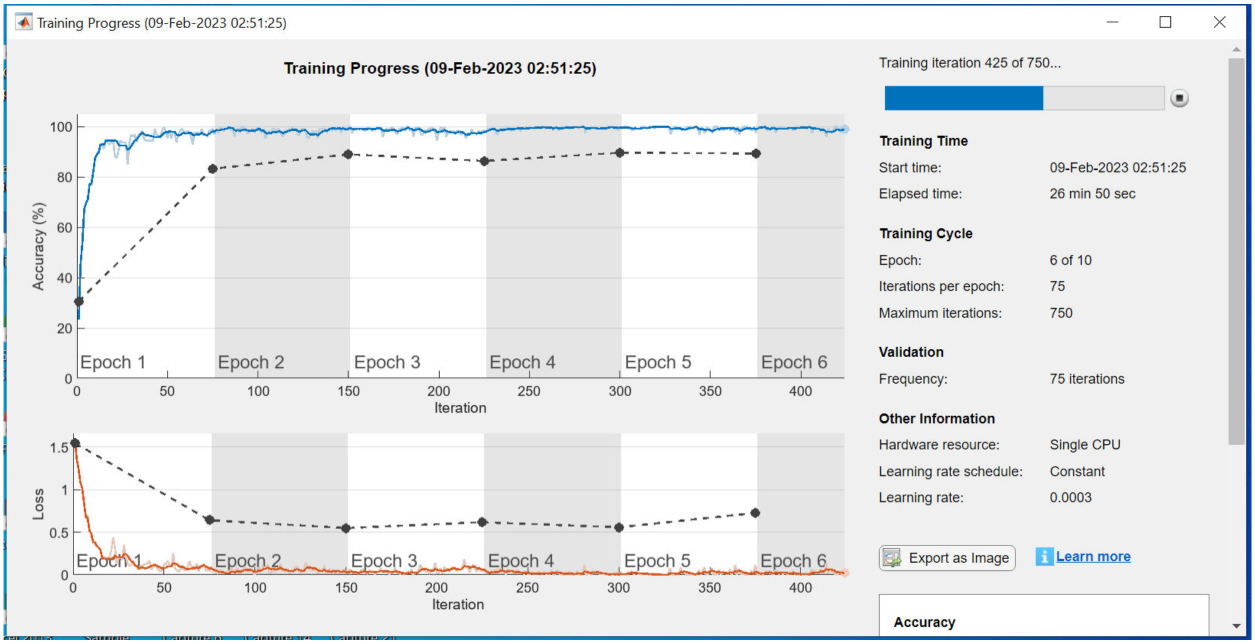
Training Time
Start time: 09-Feb-2023 02:51:25
Elapsed time: 15 min 3 sec

Training Cycle
Epoch: 3 of 10
Iterations per epoch: 75
Maximum iterations: 750

Validation
Frequency: 75 iterations

Other Information
Hardware resource: Single CPU
Learning rate schedule: Constant
Learning rate: 0.0003

Figure 6: Training Progressive Graph for the Developed CNN Model at Iteration 225 of 750.


Training Progress (09-Feb-2023 02:51:25)

Training Progress (09-Feb-2023 02:51:25)

Training iteration 425 of 750...

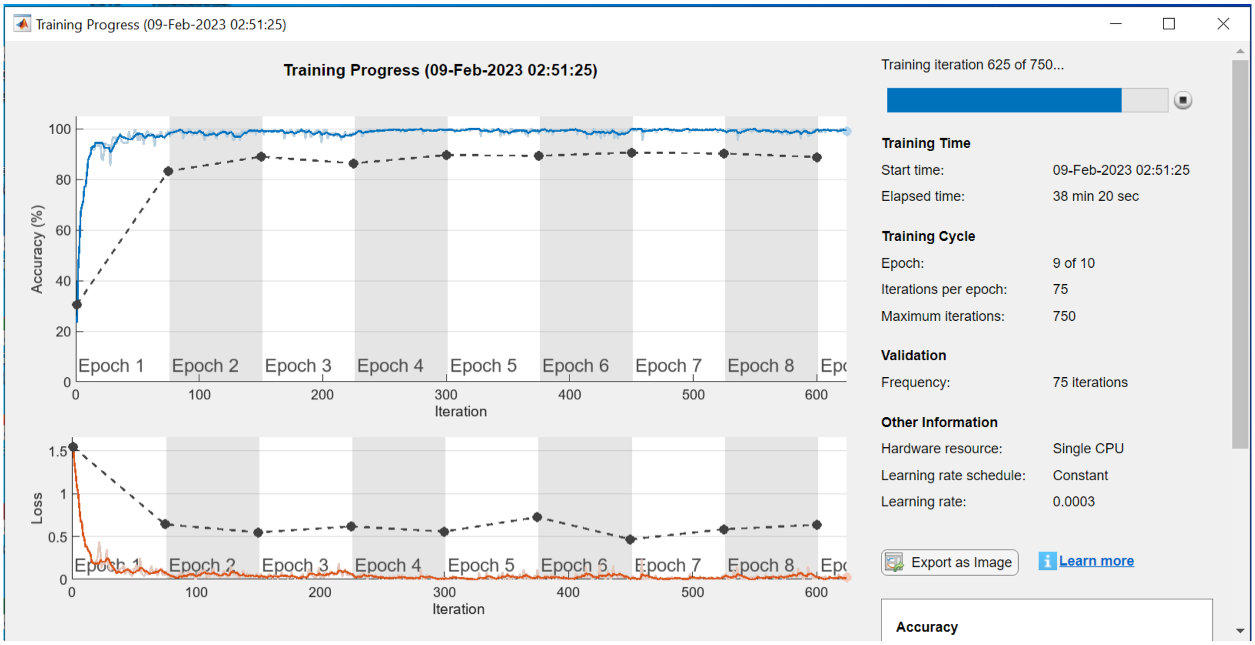
Training Time
Start time: 09-Feb-2023 02:51:25
Elapsed time: 26 min 50 sec

Training Cycle
Epoch: 6 of 10
Iterations per epoch: 75
Maximum iterations: 750

Validation
Frequency: 75 iterations

Other Information
Hardware resource: Single CPU
Learning rate schedule: Constant
Learning rate: 0.0003

Figure 7: Training Progressive Graph for the Developed CNN Model at Iteration 425 of 750.


Training Progress (09-Feb-2023 02:51:25)

Training iteration 625 of 750...

Training Time

Start time: 09-Feb-2023 02:51:25

Elapsed time: 38 min 20 sec

Training Cycle

Epoch: 9 of 10

Iterations per epoch: 75

Maximum iterations: 750

Validation

Frequency: 75 iterations

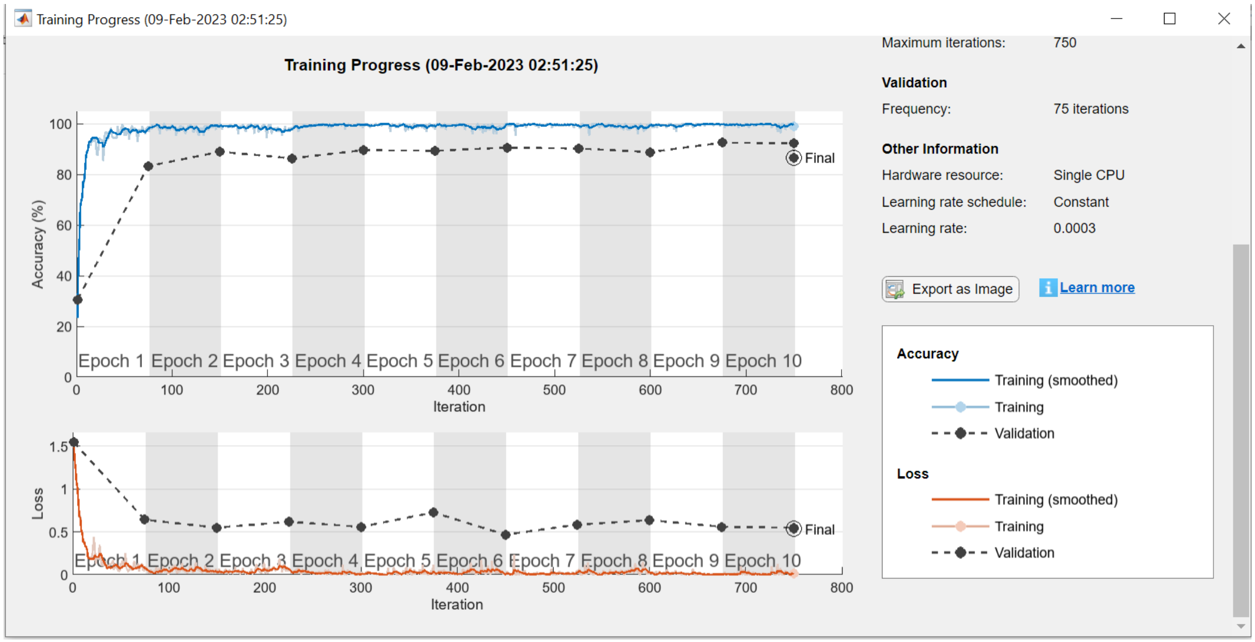
Other Information

Hardware resource: Single CPU

Learning rate schedule: Constant

Learning rate: 0.0003

Figure 8: Training Progressive Graph for the Developed CNN Model at Iteration 625 of 750.


Training Progress (09-Feb-2023 02:51:25)

Maximum iterations: 750

Validation

Frequency: 75 iterations

Other Information

Hardware resource: Single CPU

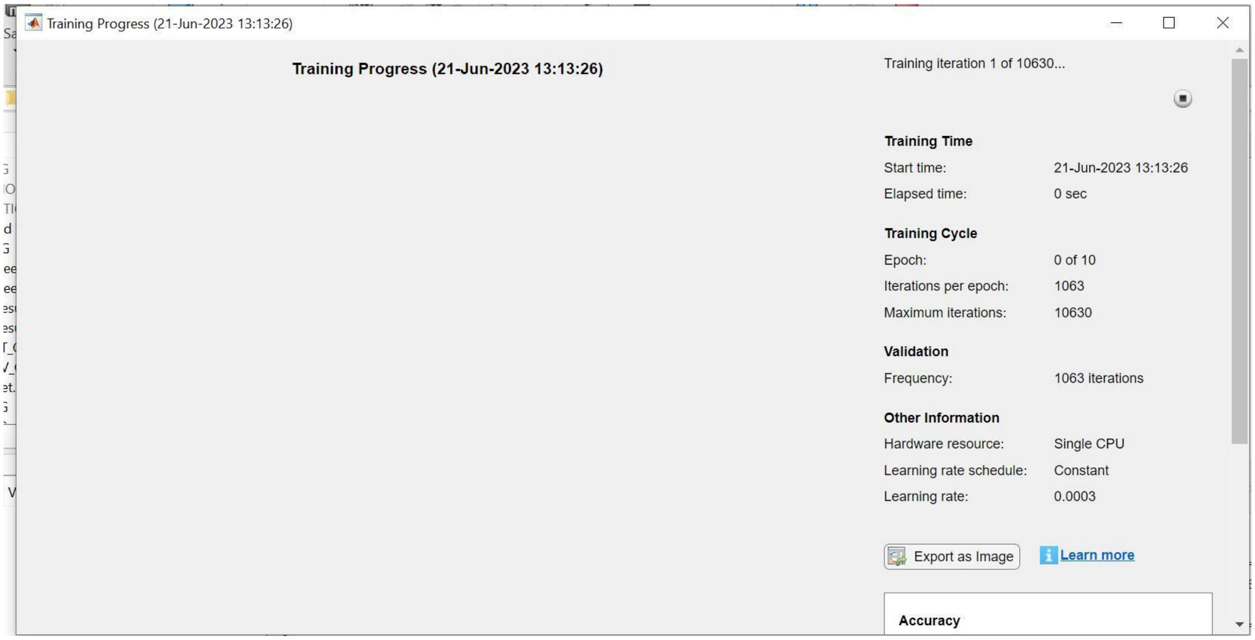
Learning rate schedule: Constant

Learning rate: 0.0003

Figure 9: Training Progressive Graph for the Dveloped CNN Model at Iteration 750 of 750.

2. *Results of the Network Training Section for Dataset B.*

70% (980 samples) of the datasets of each dialect class were used in training the network. The speech signals obtained were classified into five dialect classes. Figures 10 to 14 show the progressive training graphs of 10,630 iterations with maximum epoch of 10, for the classification of the dialects classes.



Training Progress (21-Jun-2023 13:13:26)

Training Progress (21-Jun-2023 13:13:26)

Training iteration 1 of 10630...

Training Time

Start time:	21-Jun-2023 13:13:26
Elapsed time:	0 sec

Training Cycle

Epoch:	0 of 10
Iterations per epoch:	1063
Maximum iterations:	10630

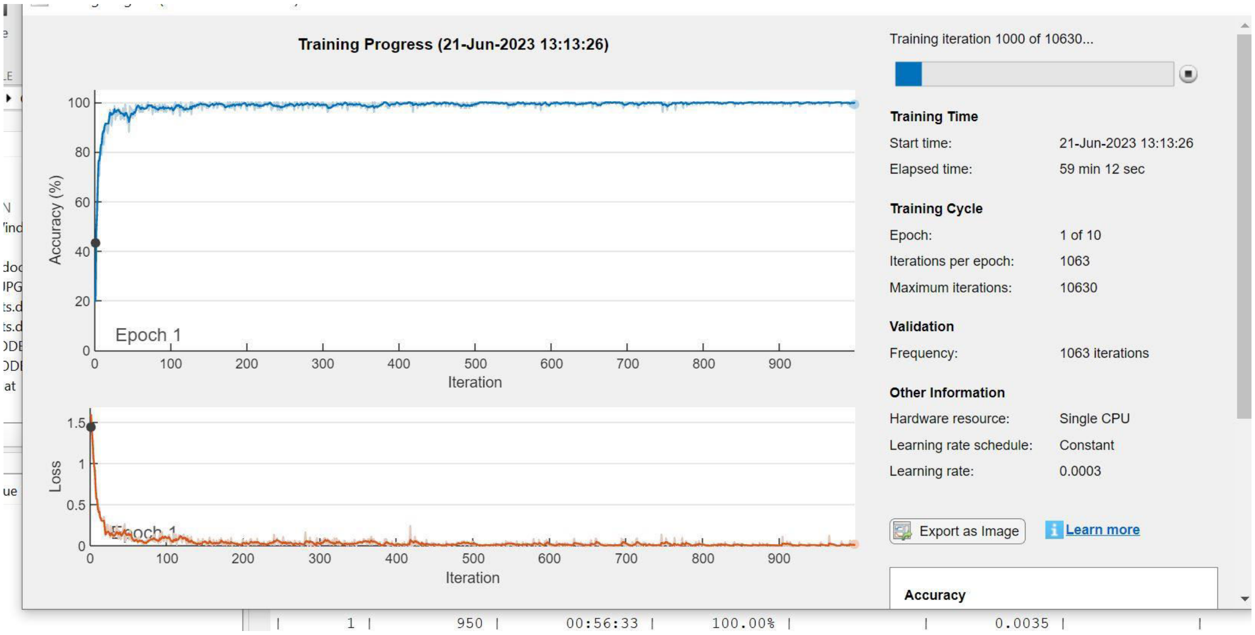
Validation

Frequency:	1063 iterations

Other Information

Hardware resource:	Single CPU
Learning rate schedule:	Constant
Learning rate:	0.0003

Figure 10: Training Progressive Graph for the Dveloped CNN Model at Iteration 1 of 10,630.



Training iteration 1000 of 10630...

Training Time

Start time: 21-Jun-2023 13:13:26

Elapsed time: 59 min 12 sec

Training Cycle

Epoch: 1 of 10

Iterations per epoch: 1063

Maximum iterations: 10630

Validation

Frequency: 1063 iterations

Other Information

Hardware resource: Single CPU

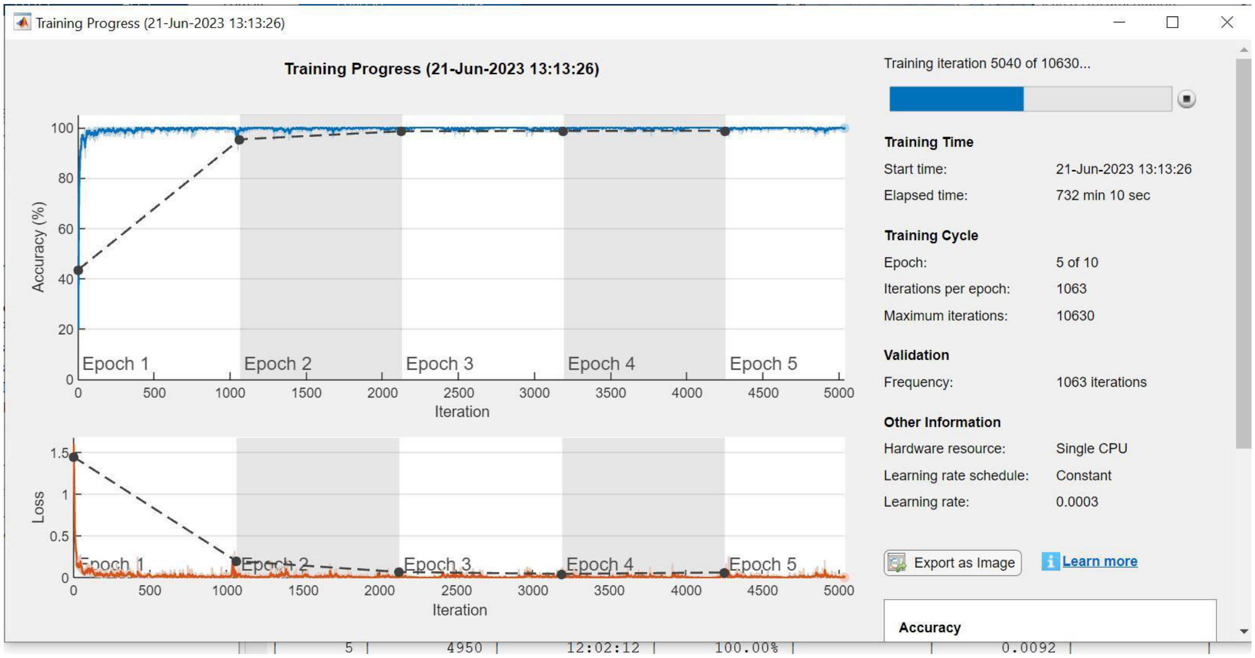
Learning rate schedule: Constant

Learning rate: 0.0003

Accuracy

```
|        1 |        950 |      00:56:33 |       100.00% |            |      0.0035 |            |
```

Figure 11: Training Progressive Graph for the Developed CNN Model at Iteration 1,000 of 10,630.



Training iteration 5040 of 10630...

Training Time

Start time: 21-Jun-2023 13:13:26

Elapsed time: 732 min 10 sec

Training Cycle

Epoch: 5 of 10

Iterations per epoch: 1063

Maximum iterations: 10630

Validation

Frequency: 1063 iterations

Other Information

Hardware resource: Single CPU

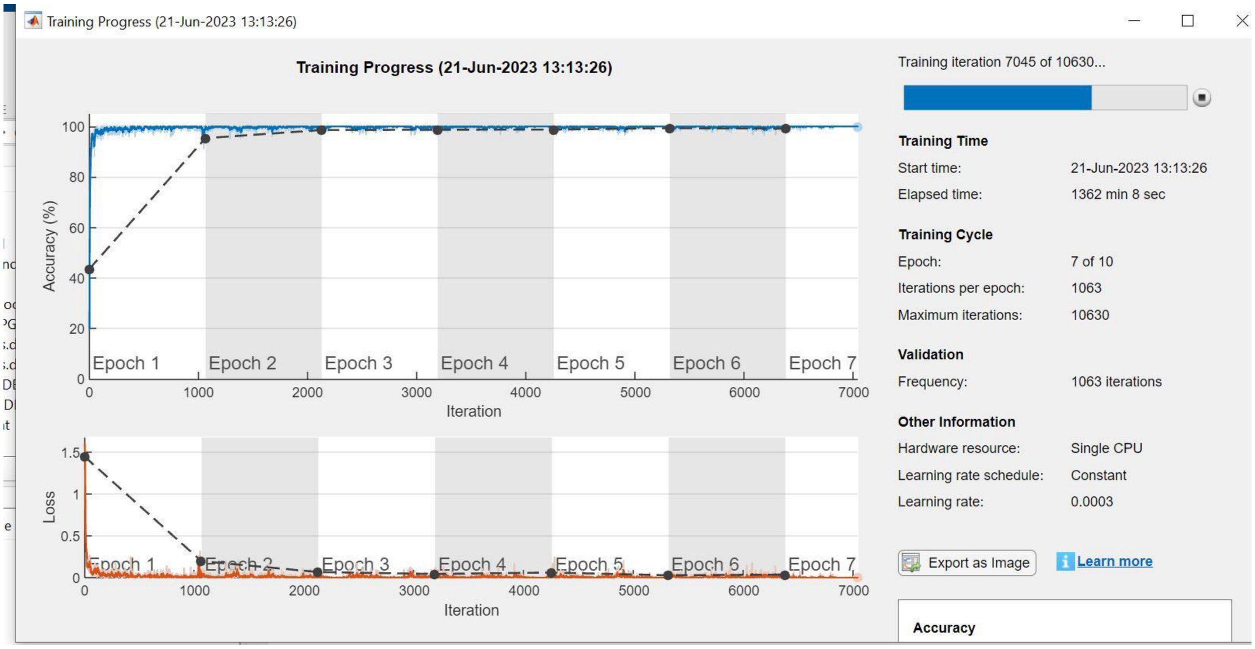
Learning rate schedule: Constant

Learning rate: 0.0003

Accuracy

```
|        5 |       4950 |      12:02:12 |       100.00% |            |      0.0092 |            |
```

Figure 12: Training Progressive Graph for the Developed CNN Model at Iteration 5,040 of10,630.



Training Progress (21-Jun-2023 13:13:26)

Training Progress (21-Jun-2023 13:13:26)

Training iteration 7045 of 10630...

Training Time

Start time:	21-Jun-2023 13:13:26
Elapsed time:	1362 min 8 sec

Training Cycle

Epoch:	7 of 10
Iterations per epoch:	1063
Maximum iterations:	10630

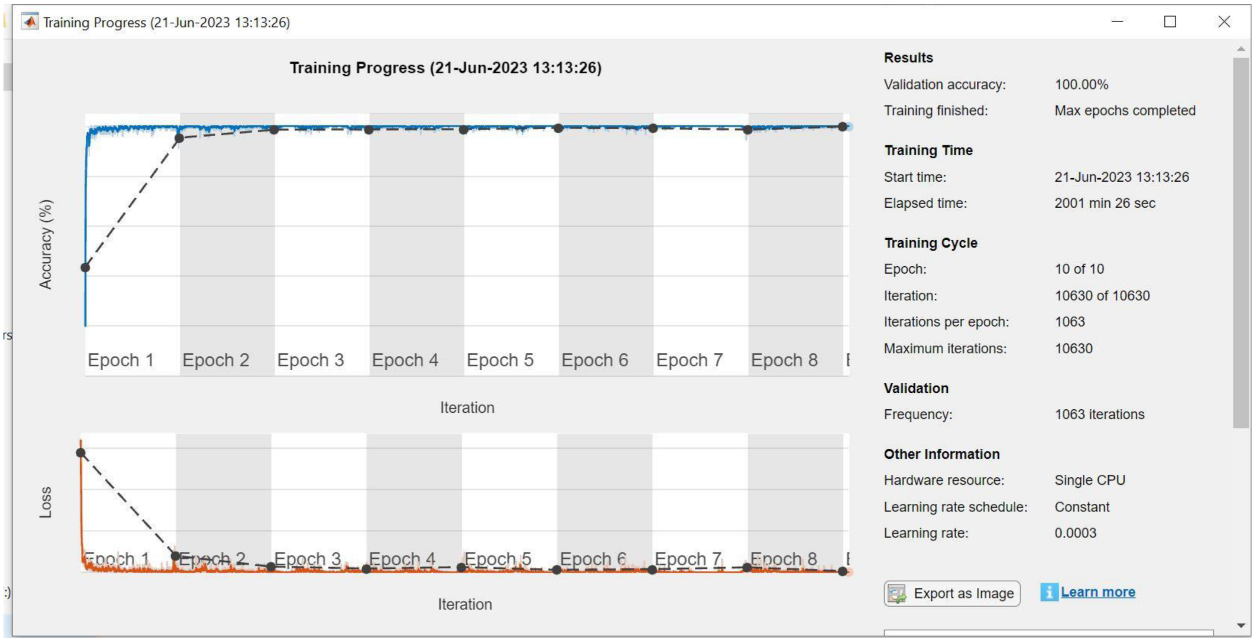
Validation

Frequency:	1063 iterations

Other Information

Hardware resource:	Single CPU
Learning rate schedule:	Constant
Learning rate:	0.0003

Figure 13: Training Progressive Graph for the Developed CNN Model at Iteration 7,045 of 10,630.



Training Progress (21-Jun-2023 13:13:26)

Training Progress (21-Jun-2023 13:13:26)

Results

Validation accuracy:	100.00%
Training finished:	Max epochs completed

Training Time

Start time:	21-Jun-2023 13:13:26
Elapsed time:	2001 min 26 sec

Training Cycle

Epoch:	10 of 10
Iteration:	10630 of 10630
Iterations per epoch:	1063
Maximum iterations:	10630

Validation

Frequency:	1063 iterations

Other Information

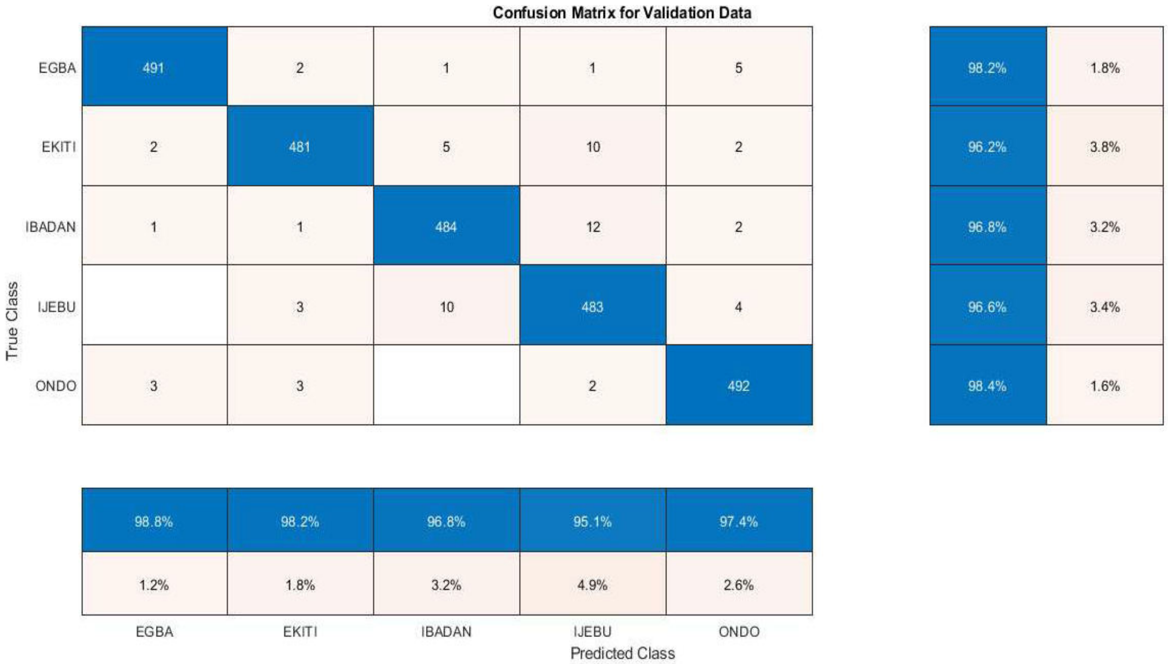
Hardware resource:	Single CPU
Learning rate schedule:	Constant
Learning rate:	0.0003

Figure 14: Training Progressive Graph for the Developed CNN Model at Iteration 10,630 of

10,630.

3. Results of the Network Validation Data and Predicted Class for Datasets A and B.

For data set A, 20% (20 samples) each of the data set were used for network validation. Figure 15shows the Confusion Matrix for Validation Data and Predicted Class while 20% (280 samples) each of the dataset were used for data set B (Figure 16).



Confusion Matrix for Validation Data

True Class	EGBA	EKITI	IBADAN	IJEBU	ONDO		
EGBA	491	2	1	1	5	98.2%	1.8%
EKITI	2	481	5	10	2	96.2%	3.8%
IBADAN	1	1	484	12	2	96.8%	3.2%
IJEBU		3	10	483	4	96.6%	3.4%
ONDO	3	3		2	492	98.4%	1.6%
	98.8%	98.2%	96.8%	95.1%	97.4%		
	1.2%	1.8%	3.2%	4.9%	2.6%		

Figure 15: Confution Matrix for Validation Data/ Predicted Class.

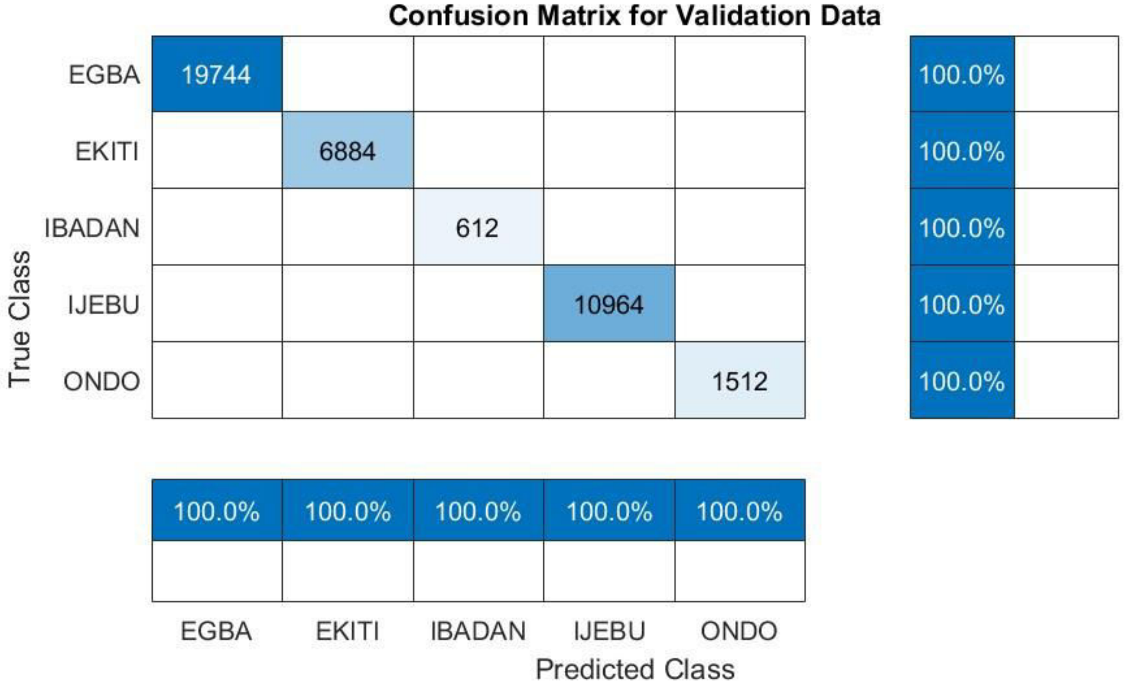


Figure 16: Confution Matrix for Validation Data/ Predicted Class

4. Results of Network Prediction.

10% of the datasets (10 samples) of each dialect class were used for dialects prediction. Table 1 shows the Confusion Matrix of the speech signals predicted for dataset A.

Table1 Confusion matrix of the dialects predicted for dataset A.

				PREDICTED		
	ACTUAL	**EGBA**	**EKITI**	**IBADAN**	**IJEBU**	**ONDO**
	EGBA	10	0	0	0	0
	EKITI	0	10	0	0	0
	IBADAN	0	1	9	0	0
	IJEBU	1	0	0	9	0
	ONDO	0	0	1	0	9

5. Performance Evaluation of the Developed Model Using Dataset A.

Considering Table 1, the performance evaluation of the developed model is determined.

Total number of samples for a class is the sum of the corresponding row TP +FN

FN for a class = sum of the corresponding rows excluding TP

FP = sum of corresponding column excluding TP

TN = sum of all columns and rows excluding that class column and row.

From table 1;

Sum of all columns and rows = 100

EGBA:

TP + FN = 10, TP = 10, FN = 0, FP = 1, TN = 39

$$Accuracy = \frac{\text{TP+TN}}{\text{Total Sample}} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\frac{10+39}{50} = \frac{49}{50} \text{ x 100\% = 98\%} \qquad (2)$$

IJEBU:

TP + FN = 10, TP = 9, FN = 1, FP = 0, TN = 39

$$Accuracy = \frac{\text{TP+TN}}{\text{TOTAL}} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+39}{50}\text{x100 = 96\%} \qquad (3)$$

ONDO:

TP + FN = 10, TP = 9, FN = 1, FP = 0 and TN = 40

$$Accuracy = \frac{\text{TP+TN}}{\text{TOTAL}} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+40}{50} \text{ x 100\% = 98\%} \qquad (4)$$

EKITI:

TP + FN = 10z, TP = 10, FN = 0, FP = 1, TN = 39

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{total samples}} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{10+39}{50} \text{X 100\% = 98\%} \qquad (5)$$

IBADAN:

TP + FN = 10, TP = 9, FN = 1, FP = 1, TN = 39

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{total samples}} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+39}{50} \text{ x 100\% = 96\%} \qquad (6)$$

IV. Discussion

The effects of datasets size was investigated in this work. The audio samples of Egba, Ekiti, Ibadan, Ijebu and Ondo dialects were collected from participants via mobile phones, radio and sound recorders. Two (2) datasets A and B were used. Dataset A has a total number of 500 samples (100 samples for each of the classes). Dataset B has a total number of 7000 samples (1400 samples for each of the classes). Both datasets were divided into 70% for training the network, 20% for validation and 10% for prediction. The datasets were first converted to ".wav" format for efficient training using CNN. These

audio waveforms were later converted to auditory-based spectrograms (see Figures 2 and 3). Figure 4 shows the block diagram of the CNN-based classification Model process.
Figures 5 to 9 show, 750 iterations of the progressive training graphs during network training for dataset A. Figures 10 to 14 display 10,630 iterations of the progressive training graphs during network training of dataset B.The upper part of the graphs showed accuracy against iteration while the lower parts showed loss against iteration. Figures 15 and 16 show the Confusion Matrices for validation data for Datasets A and B respectively.For dataset A, the performance accuracy of the Model's predictions for Egba, Ekiti, Ibadan, Ijebu and Ondo are 98.8%, 98.2%, 96.8%, 95.1% and 97.4% respectively. For datset B, the performance accuracy of Model's prediction for the five (5) classes is 100%. This shows that the classifier performed better with large dataset, B (see Table 2).

Table 2 Comparison of experimentalpredicted class accuracy with calculated results for data sets A.

Classes	**Experimental Results for Dataset 'A' (%)**	**Evaluated Results for Dataset 'A' (%)**	**Experimental Results for Dataset 'B' (%)**
EGBA	98.80	98.00	100.00
EKITI	98.20	98.00	100.00
IBADAN	96.80	96.00	100.00
IJEBU	95.10	96.00	100.00
ONDO	97.40	98.00	100.00

V. Conclusion and Recommendations

This research investigated the effects of datasets sizes on the performance accuracy of dialects Classification Model. A Convolutional Neural Network (CNN) Classifier was developed.The process of achieving the objective of this research was divided into four (4) main stages namely: speech signals acquisition, data pre-processing, speech data classification and Model training/ testing and evaluation. The Model was implemented on Matlab 2022b platform. With the same Classifier, the results showed that the larger sized dataset 'B' gave a better performance accuracy that the smaller sized dataset A. however, it is recommended that the complexity of the Model be considered before increasing the datasets to avoid under-fitting in the network.

References

1. Alhanoof, A., Duaac, A., Heyam, A., Amani, S., Alanoud, B. D., Najla, A., Afnan, A.Elw. andHeba, K. (2021). Impact of dataset size on classification performance,An empirical evaluation in the medical domain. Applied sciences. Multidisciplinary digital publishing institute, 11 (2):1-18.
2. Barbedo, J.G. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153, pp 46–53.
3. Blake, C.L.; Merz, C.J. (2020). Uci repository of machine learning databases.Department of information and computer science, university of california: Irvine, CA, USA, 1998; 55.
4. Choi, Y., Lee, H. (2017).Data properties and the performance of sentiment classification for electronic commerce applications. Inf. Syst. Front. 19: 993–1012.
5. Dris, A.B.; Alzakari, N.; Kurdi, H. (2019): A systematic approach to identify an appropriate classifier for limited-sized data sets. In proceedings of the 2019 international symposium on networks, computers and communications (ISNCC), Istanbul, Turkey, 18–20 June 2019; 1- 6.
6. Linjordet, T., Balog, K. (2019). Impact of training dataset size on neural answer selection models. In proceedings of the european conference on information retrieval, cologne, germany, 14 April 2019; Springer: Cham, Switzerland, 828–835.
7. Marcoulides, G.A. (2005) Discovering knowledge in data: An introduction to data mining. Daniel T. Larose. J. Am. Stat. Assoc. 100, 1465.
8. Mehrafarin, H., Rajaee, S. and Pilehvar, M.T. (2022) On the importance of data size in probing fine-tuned models. Findings of the association for computational linguistics: ACL 2022, 228 – 238.
9. Rahman, M.S., Sultana, M. (2017): Performance of firth-and logf-type penalized methods in risk prediction for small or sparse binary data. BMC Med. Res. Methodol. 2017, 17:33.
10. Wieczorek, G., Antoniuk, I., Kurek, J., Swiderski, B., Kruk, M., Pach, J., Orłowski, A. (2019): BCT Boost segmentation with u-net in tensorflow. Mach. Graph. Vis. 2019, vol. 28: 25–34.

11. Zhu, X., Vondrick, C., Fowlkes, C.C., Ramanan, D. (2016): “Do we need more training data?” Int. J. Comput. Vis. 2016, 119: 76–92.