

## 4. RESULT ANALYSIS

---

This chapter goes through a detailed analysis of the lip extraction algorithm proposed in previous chapter, ViLiDEx. It also covers the training and testing of the model that has been created using the GJVarna Dataset. Finally, the chapter concludes based on experimental analysis.

### 4.1 TESTING OF ViLiDEx ALGORITHM

ViLiDEx algorithm is based on face landmark algorithm of Dlib. Different speakers have different utterance times for different alphabets. Hence, the number of frames is different for each speaker for each alphabet. ViLiDEx algorithm selects fixed number of key frames from total number of frames, extract ROI and save it.

Two different frame sizes are used here, 15 and 20. For understating purpose, the following section explains the working of frame extraction using ViLiDEx algorithm for frame size 20. The algorithm has been implemented using Python 3.9, Keras, Tensorflow, dlib face landmark and MobileNet model. Following is the summary of the experimental setup and configuration:

1. ViLiDEx algorithm is applied to create GJVarna dataset with two different frames sizes, 15 and 20, per speaker for 34 consonants of Gujarati language.
2. Of these 34 consonants, 15 consonants are used for alphabet recognition using CNN-LSTM based model MobileNet.

Sr. No	Parameters	Details
1	Frames	15 and 20 frames
2	Classes	(i) Guttural (ii) Palatal (iii) Retroflex (iv) Dental and (v) Labial
3	Alphabets taken	5 consonants, 3 from each class
4	Speakers	19 speakers for training and 9 speakers for testing dataset
5	Images	(i) 15 frames: (a) 4275 images for training (b) 2025 images for testing (ii) 20 frames: (a) 5700 images for training (b) 2700 images for testing

*Table 1 Training and Testing setup*

Specifications	Laptop-PC
Model	Lenovo 80Q
Processor	Intel® Core-i7-6500U CPU with 2.5 GHz
RAM	8 GB
Operating System	Windows10 Home
System Type	64-bit Operating System with the x64-based processor

Table 2 Configuration of Laptop PC used for training and testing

The following subsection explains the working of frame removal using ViLiDEx algorithm.

#### 4.1.1 Total frames =20

Frames differ = 0 (Total frames are equal to needed frames)

divisor =1

Frame numbers divisible by 1 are selected, cropped and saved.

#### 4.1.2 Total frames are multiple of 20. (40, 60, 80...)

divisor = total frames / 2

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20

Figure 1 Frame selection for total frames 20

**Example 1: Total frames = 40 → divisor = 40/20 = 2**

Frame numbers divisible by 2 are selected, cropped and saved. Other frames are discarded.

See figure.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20
Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27	Frame 28	Frame 29	Frame 30
Frame 31	Frame 32	Frame 33	Frame 34	Frame 35	Frame 36	Frame 37	Frame 38	Frame 39	Frame 40

Figure 2 Frame selection when total frames are 40

**Example 2: Total frames = 60 → divisor = 60/20 = 3**

Frame numbers divisible by 3 are selected and others are discarded. See figure.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20
Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27	Frame 28	Frame 29	Frame 30
Frame 31	Frame 32	Frame 33	Frame 34	Frame 35	Frame 36	Frame 37	Frame 38	Frame 39	Frame 40
Frame 41	Frame 42	Frame 43	Frame 44	Frame 45	Frame 46	Frame 47	Frame 48	Frame 49	Frame 50
Frame 51	Frame 52	Frame 53	Frame 54	Frame 55	Frame 56	Frame 57	Frame 58	Frame 59	Frame 60

Figure 3 Frame selection when total frames are 60

#### 4.1.3 Frames are not divisible by 20

**Example 1: total frames = 30**

- Frames differ =  $30 - 20 = 10$  (10 frames should be discarded)
- Divisors from 1 to 30 [3, 5, 7, 11, 13, 17, 19, 23, 29]
- The numbers divisible by above divisors up to 30 are [3 -> 10, 5 -> 6, 7 -> 4, 11 -> 2, 13 -> 2, 17 -> 1, 19 -> 1, 23 -> 1, 29 -> 1)
- Divisor 3 is selected as numbers divisible by 3 equal to 10.
- Frames divisible by 3 are discarded and others are saved.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20
Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27	Frame 28	Frame 29	Frame 30

Figure 4 Frame selection when total frames are 30

**Example 2. Total frames = 24**

- Frames differ =  $24 - 20 = 4$  (4 frames should be discarded)
- Divisors from 1 to 24 are [3, 5, 7, 11, 13, 17, 19, 23].
- Numbers divisible by the above divisors up to 24 are [3 -> 8, 5 -> 4, 7 -> 3, 11 -> 2, 13 -> 1, 17 -> 1, 19 -> 1, 23 -> 1]
- Divisor 5 is selected as numbers divisible by 5 is 4 and that match with frames differ
- Frames divisible by 5 will be discarded and others are saved. See figure.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8
Frame 9	Frame 10	Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16
Frame 17	Frame 18	Frame 19	Frame 20	Frame 21	Frame 22	Frame 23	Frame 24

Figure 5 Frame selection when total frames are 24

**Example 3 total frames = 27**

- Frames differ = 7
- Total 8 divisors from 1 to 27 are [3, 5, 7, 11, 13, 17, 19, 23]
- Numbers divisible by above divisors up to 27 are [9, 5, 3, 2, 2, 1, 1, 1]
- Divisor 3 is discarded as numbers divisible by 3 are  $9 > 7$  (frames differ)
- Divisors 5 and 7 are not eligible as numbers divisible by 5 and 7 are  $5 + 3 = 8$ , now subtract common numbers divisible by 5 and 7, which makes total number  $8 - 0 = 8 > 7$  (frames differ)
- Divisors 5 and 11 are selected as numbers divisible by 5 and 11 are  $5 + 2 = 7 - 0$  (common numbers of 5 and 11 up to 27)
- Frames divisible by 5 and 11 are discarded, others are saved. See figure.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9
Frame 10	Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18
Frame 19	Frame 20	Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27

Figure 6 Frame selection when total frames are 27

**Example 4:** Total frames = 39

- Frames differ = 19
- Total 11 divisors from 1 to 39 are [3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37]
- Numbers divisible by above divisors up to 39 are [13, 7, 5, 3, 3, 2, 2, 1, 1, 1, 1]
- Total numbers for [3, 5] are  $13 + 7 - 2$  (common numbers of 3 and 5) = 18
- If 7 is selected total numbers are  $18 + 5 = 23 - 1$  (common number of 3 and 7 -> 21) - 1 (common number of 7 and 5 -> 35) = 21 > 19 → 7 is not selected.
- If 11 is selected total numbers are  $18 + 3 = 21 - 1$  (common number of 3 and 11) - 0 (common number of 5 and 11) = 20 > 19 → 11 is not selected.
- If 13 is selected total numbers are  $18 + 3 = 21 - 1 = 20 > 19$  → 13 is not selected.
- For 17 and 19 total numbers are  $20 > 19$  → 17 or 19 are not selected.
- For 23 total numbers are  $18 + 1 = 19 - 0 = 19$  → 23 is selected.
- Divisor [3, 5, 23] are selected.
- Frames divisible by 3, 5 and 23 are discarded and others are saved.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20
Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27	Frame 28	Frame 29	Frame 30
Frame 31	Frame 32	Frame 33	Frame 34	Frame 35	Frame 36	Frame 37	Frame 38	Frame 39	

Figure 7 Frame selection when total frames are 39

**Example 5:** total frames = 50

- Frames differ = 30
- Total divisors are [3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47]
- Numbers divisible by above divisors are [16, 10, 7, 4, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1]
- For divisors [3,5], total numbers are  $16 + 10 = 26 - 3 = 23$
- If 7 is selected  $23 + 7 = 30 - 3 = 27$
- If 11 is selected  $27 + 4 = 31 - 1 = 30$
- Divisor [3, 5, 7, 11] are selected.
- Frames divisible by 3, 5, 7, and 11 are discarded, others are saved.

Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Frame 11	Frame 12	Frame 13	Frame 14	Frame 15	Frame 16	Frame 17	Frame 18	Frame 19	Frame 20
Frame 21	Frame 22	Frame 23	Frame 24	Frame 25	Frame 26	Frame 27	Frame 28	Frame 29	Frame 30
Frame 31	Frame 32	Frame 33	Frame 34	Frame 35	Frame 36	Frame 37	Frame 38	Frame 39	Frame 40
Frame 41	Frame 42	Frame 43	Frame 44	Frame 45	Frame 46	Frame 47	Frame 48	Frame 49	Frame 50

Figure 8 Frame selection when total frames 50

## 4.2 ALPHABET RECOGNITION

Total number of frames per alphabet affects the recognition rate. The subset of GJVarna dataset is created for frame sizes 15, 20. Three alphabets from each class were taken and passed through ViLiDEx algorithm for 28 speakers. 19 speakers used for training and 9 speakers for testing. With the combination of different values of learning rate and validation ratio four cases are used. Parameters, cases, and Results are shown in table.

Class	Alphabets
Guttural	କ (ka), ଲ (kha), ଗ (ga)
Palatal	ଚ (cha), ଛ (chha), ଜ (ja)
Retroflex	ଟ (tta), ଠ (thha), ଡ (dda)
Dental	ଡ (ta), ଥ (tha), ଡ (da)
Labial	ପ (pa), ଫ (pha), ବ (ba)

Table 3 Class wise alphabet

A total of 4275 and 5700 image have been taken for training purpose for 15 and 20 frames respectively. For 15 frames, 855 (20%) and 1070 (25%) of input images have been taken for validation. Similarly, for 20 frames, 1140 (20%) and 1425 (25%) of input images have been taken for validation. For 15 frames, 27 speakers and 5 classes, a total of 2025 images have been taken for testing. For 20 frames, 27 speakers and 5 classes, a total of 2700 images have been taken for testing.

Hence, based on above parameters, four different cases have been considered for 15 and 20 frames with reference to learning rate and validation ratio. Epochs have been taken from 5 to 50 for each case. Following section shows the result analysis in terms of training, validation and testing accuracy followed by confusion matrices of four different cases mentioned below.

Case 1		Case 2		Case 3		Case 4	
LR	VR	LR	VR	LR	VR	LR	VR
0.0001	0.2	0.0002	0.2	0.0001	0.25	0.0002	0.25

Table 4 Four cases for result analysis

### 4.2.1 Training, Validation and Testing

#### 4.2.1.1 Results for 15 frames

LR	Epochs	Train Accuracy	Val Accuracy	Test Accuracy	LR	Epochs	Train Accuracy	Val Accuracy	Test Accuracy
<b>Case 1 0.0001</b>	5	0.27	0.19	0.2	<b>Case 3 0.0001</b>	5	0.27	0.26	0.27
	10	0.29	0.25	0.25		10	0.32	0.27	0.27
	15	0.36	0.16	0.28		15	0.57	0.27	0.27
	20	0.5	0.25	0.27		20	0.68	0.3	0.31
	25	0.6	0.25	0.27		25	0.73	0.38	0.33
	30	0.76	0.32	0.36		30	0.78	0.33	0.33
	35	0.78	0.32	0.32		35	0.78	0.38	0.33
	40	0.77	0.32	0.3		40	0.78	0.42	0.38
	50	0.78	0.32	0.31		50	0.78	0.43	0.37
<b>Average</b>		56.78%	26.44%	28.44%			63.22%	33.78%	31.78%
<b>Case 2 0.0002</b>	5	0.26	0.19	0.24	<b>Case 4 0.0002</b>	5	0.26	0.2	0.22
	10	0.32	0.16	0.29		10	0.35	0.21	0.26
	15	0.33	0.24	0.28		15	0.38	0.26	0.27
	20	0.36	0.19	0.27		20	0.56	0.25	0.3
	25	0.43	0.23	0.24		25	0.6	0.3	0.3
	30	0.72	0.19	0.24		30	0.66	0.28	0.3
	35	0.8	0.26	0.29		35	0.72	0.25	0.25
	40	0.82	0.28	0.28		40	0.76	0.36	0.32
	50	0.82	0.3	0.28		50	0.76	0.38	0.33
<b>Average</b>		54.00%	22.67%	26.78%			56.11%	27.67%	28.33%

Table 5 Four cases for result analysis (15 frames)

It has been noted that accuracy is increasing as the number of epochs is increasing. From 5 to 25 epochs, the training, validation and testing accuracy starts increasing. However, between 30 to 50 epochs the accuracy remains almost steady and doesn't increase. The learning rate 0.0001

gives better Accuracy than 0.0002. At the same time the validation ratio 0.25 gives better results than 0.2. Another point to be noted here is that the larger the Dataset is, with repeated speakers, the better the results. The following graph shows the summarized performance of all four cases for 15 frames.

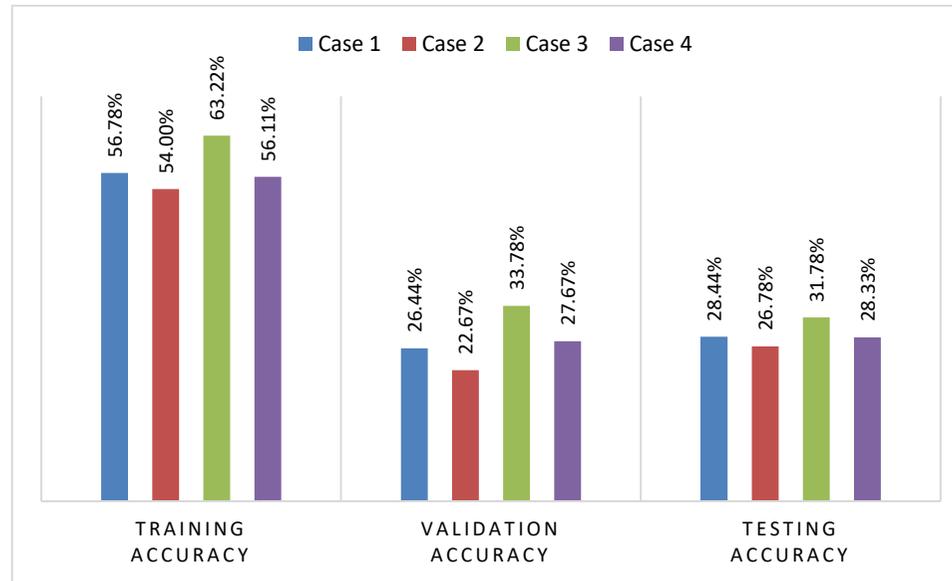


Figure 9 Accuracy comparison for 15 frames

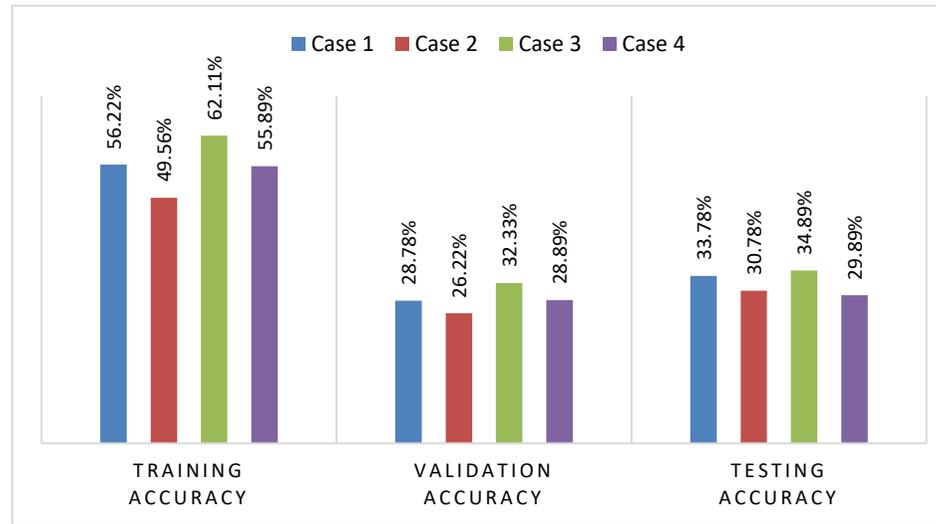
#### 4.2.1.2 Results for 20 frames

LR	Epochs	Train Accuracy	Val Accuracy	Test Accuracy	LR	Epochs	Train Accuracy	Val Accuracy	Test Accuracy
<b>Case 1 0.0001</b>	5	0.24	0.18	0.27	<b>Case 3 0.0001</b>	5	0.32	0.25	0.25
	10	0.28	0.24	0.25		10	0.42	0.27	0.32
	15	0.43	0.3	0.33		15	0.53	0.25	0.33
	20	0.48	0.27	0.35		20	0.64	0.26	0.32
	25	0.57	0.32	0.38		25	0.71	0.32	0.36
	30	0.72	0.33	0.34		30	0.73	0.37	0.38
	35	0.76	0.3	0.37		35	0.73	0.4	0.4
	40	0.78	0.33	0.37		40	0.75	0.39	0.39
	50	0.8	0.32	0.38		50	0.76	0.4	0.39
<b>Average</b>		56.22%	28.78%	33.78%			62.11%	32.33%	34.89%
<b>Case 2 0.0002</b>	5	0.18	0.14	0.19	<b>Case 4 0.0002</b>	5	0.24	0.15	0.24
	10	0.22	0.19	0.22		10	0.36	0.25	0.28
	15	0.34	0.24	0.25		15	0.39	0.29	0.31
	20	0.4	0.21	0.27		20	0.51	0.26	0.3
	25	0.46	0.26	0.37		25	0.6	0.3	0.25
	30	0.59	0.33	0.35		30	0.63	0.31	0.3
	35	0.7	0.33	0.35		35	0.74	0.33	0.31
	40	0.77	0.32	0.39		40	0.77	0.36	0.36
	50	0.8	0.34	0.38		50	0.79	0.35	0.34
<b>Average</b>		49.56%	26.22%	30.78%			55.89%	28.89%	29.89%

Table 6 Four cases for result analysis

It has similar observations as 15 frames. The accuracy is increasing as the number of epochs is increasing. From 5 to 25 epochs, the training, validation and testing accuracy starts increasing. However, between 30 to 50 epochs the accuracy remains almost steady and doesn't increase.

Here too, the learning rate 0.0001 gives better Accuracy than 0.0002 whereas validation ratio 0.25 gives better results than 0.2. It follows the same conclusion, the larger the dataset is, with repeated speakers, the better the results are. The following graph shows the summarized performance of all four cases for 20 frames.



*Figure 10 Accuracy comparison for 20 frames*

## 4.2.2 Precision, Recall and F1 Score Analysis

The following subsections show the observations based on the confusion matrices generated for 15 and 20 different frames with epochs ranging from 15 to 40 and for all five categories of alphabet.

### 4.2.2.1 For 15 frames

#### 1. Case 1 for LR 0.0001 and VR 0.2

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	100%	4%	7%	27%	41%	32%	35%	26%	30%	23%	48%	31%	35%	22%	27%
20	21%	37%	27%	45%	48%	46%	23%	11%	15%	23%	11%	15%	21%	26%	23%
25	32%	26%	29%	60%	33%	43%	25%	7%	11%	5%	4%	4%	24%	63%	35%
30	44%	26%	33%	55%	41%	47%	32%	26%	29%	19%	15%	17%	36%	74%	48%
35	30%	22%	26%	53%	37%	43%	10%	4%	5%	-	4%	-	30%	96%	46%
40	24%	37%	29%	32%	37%	34%	35%	30%	32%	16%	11%	13%	45%	33%	38%

Table 7 Precision, Recall, and F1 score for Case 1(15 frames)

Results for Guttural, Retroflex, and Dental are fluctuating for different epochs. However, for Palatal and Labial classes, it is not changing for different epoch values. For 40 epochs, the precision, recall and F1 score remain balanced. The following graph shows the summarized performance of all the classes for LR 0.0001 and validation ration 0.2.

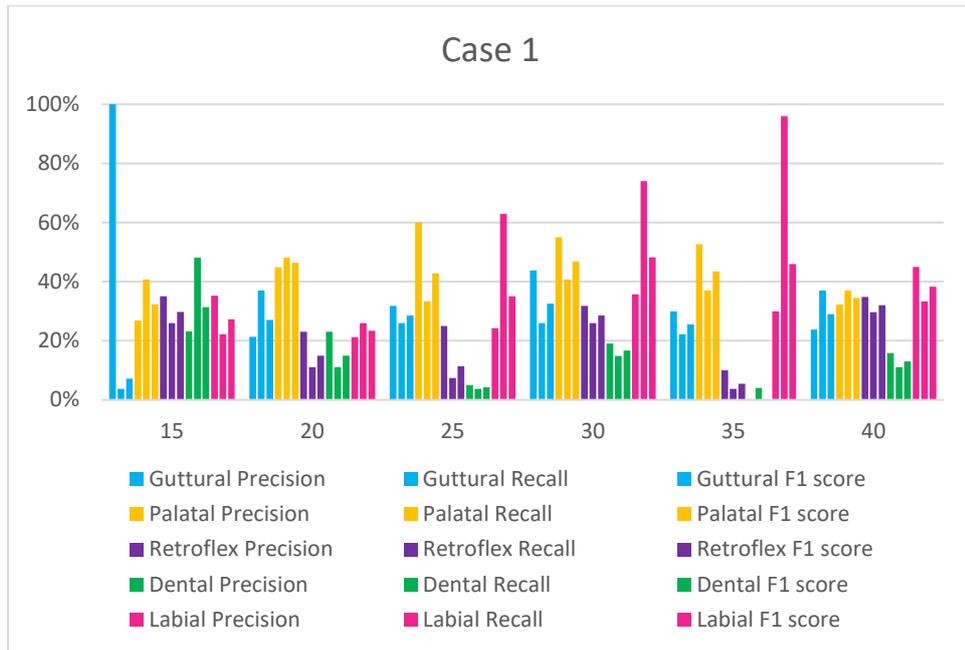


Figure 11 Accuracy comparison for different classes for case 1 (15 frames)

2. Case 2 for LR 0.0002 and VR 0.2

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	22%	74%	33%	50%	15%	23%	40%	7%	13%	16%	15%	15%	75%	11%	19%
20	23%	56%	33%	83%	84%	84%	24%	26%	25%	30%	11%	16%	100%	7%	14%
25	40%	7%	13%	42%	30%	35%	13%	15%	14%	25%	63%	36%	25%	11%	15%
30	23%	30%	26%	24%	44%	31%	20%	19%	19%	24%	19%	21%	33%	4%	7%
35	30%	48%	37%	26%	41%	32%	30%	11%	16%	13%	11%	12%	35%	22%	27%
40	26%	30%	28%	29%	26%	27%	24%	44%	31%	23%	11%	15%	44%	26%	33%

Table 8 Precision, Recall, and F1 score for case 2 (15 frames)

The table shows that the precision, recall and F1 score for all classes is not balanced for epochs 15, 20, 25, and 30 which gets balanced at 35 and 40 epochs. The following graph shows the performance summary for case 2.

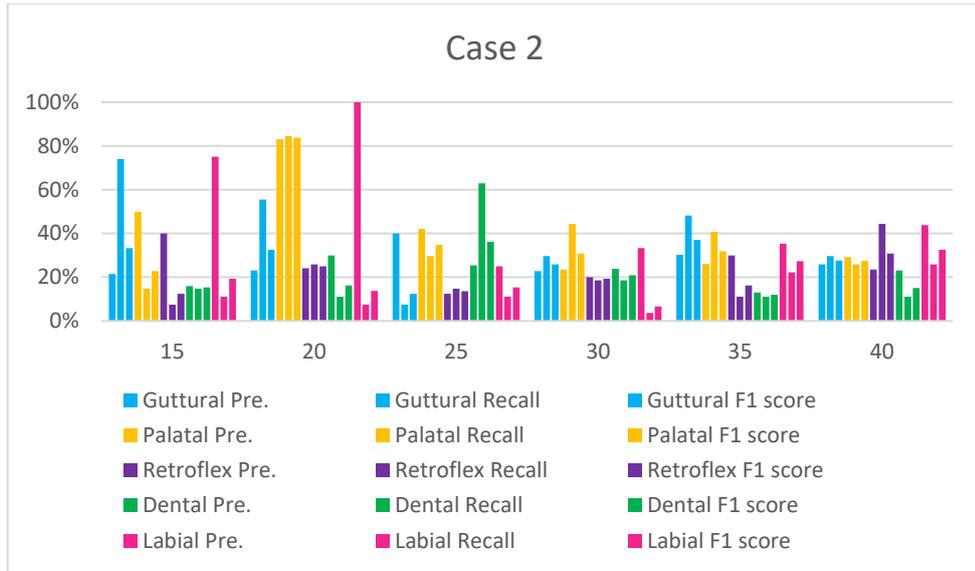


Figure 12 Accuracy comparison for different classes for case 2 (15 frames)

### 3. Case 3 for LR 0.0001 and VR 0.25

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	29%	70%	41%	42%	19%	26%	36%	19%	24%	14%	15%	14%	43%	22%	29%
20	38%	19%	25%	52%	48%	50%	38%	30%	33%	17%	26%	21%	37%	48%	42%
25	34%	41%	37%	37%	52%	43%	26%	33%	30%	13%	4%	6%	39%	33%	36%
30	43%	22%	29%	54%	48%	51%	24%	30%	26%	20%	33%	25%	41%	26%	32%
35	30%	11%	16%	43%	37%	40%	17%	15%	16%	21%	33%	26%	41%	56%	47%
40	40%	22%	29%	41%	59%	48%	27%	33%	30%	21%	11%	15%	47%	59%	52%

Table 9 Precision, Recall and F1 score for Case 3 (15 frames)

The precision, recall and F1 score for Dental class fluctuates for different epochs and are stable and balanced for other classes for all epochs. Overall performance is good for all classes at 30 epochs. The following graph shows the performance of case 3 for all classes with various epochs.

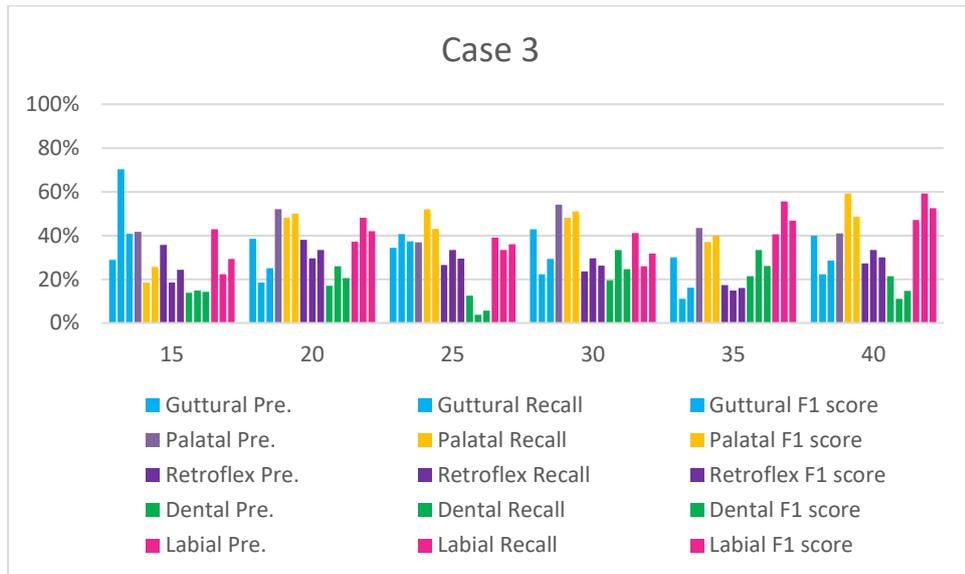


Figure 13 Accuracy comparison for different class for case 3 (15 frames)

4. Case 4 for LR 0.0002 and VR 0.25

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	25%	7%	11%	38%	22%	28%	100%	4%	7%	21%	59%	30%	28%	33%	31%
20	34%	41%	37%	29%	85%	43%	100%	11%	20%	38%	22%	28%	50%	7%	13%
25	24%	26%	25%	42%	48%	45%	17%	4%	6%	22%	7%	11%	32%	70%	44%
30	26%	22%	24%	30%	59%	40%	14%	7%	10%	17%	7%	10%	41%	48%	44%
35	25%	44%	32%	25%	33%	29%	27%	11%	16%	17%	19%	18%	40%	15%	22%
40	44%	30%	36%	27%	48%	35%	31%	33%	32%	19%	15%	17%	47%	33%	39%

Table 10 Precision, Recall, and F1 score for Case 4 (15 frames)

The performance of Palatal class is balanced for all epochs. However, it is not balanced for other classes up to 30 epochs. For example, for Guttural, with 15 epochs it is imbalanced. Similarly, for Retroflex, Dental and Labial, it is highly imbalanced for 15 to 30 epochs. It is balanced at 35 and 40 epochs. The following is the graphical representation of the precision, recall and F1 score discussed.

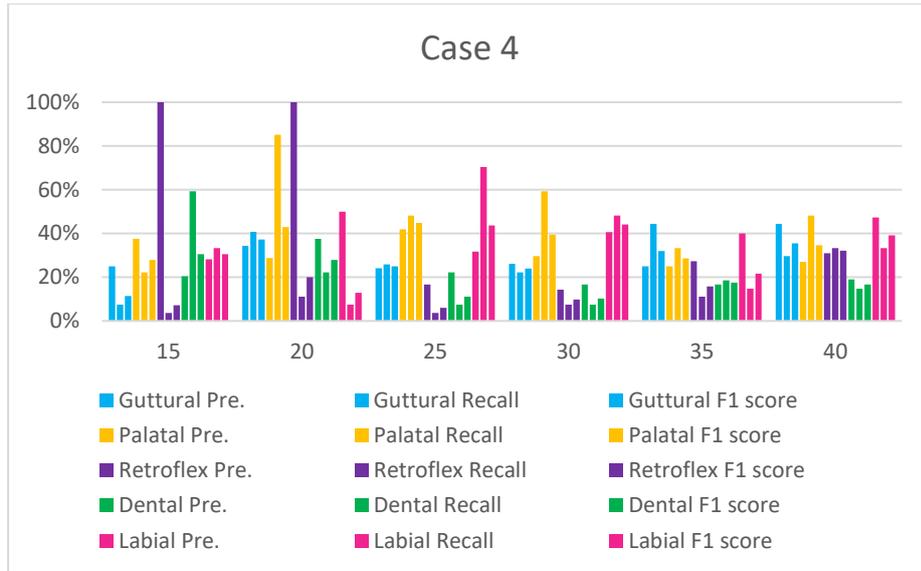


Figure 14 Accuracy comparison for different class for case 4 (15 frames)

Overall accuracy is good for 40 epochs, for all cases. Best accuracy is obtained for case 3 with LR 0.0001 and VR 0.25

Epochs	Case 1	Case 2	Case 3	Case 4
15	28%	24%	29%	25%
20	27%	37%	34%	33%
25	27%	25%	33%	31%
30	36%	23%	32%	29%
35	33%	27%	30%	24%
40	30%	27%	37%	32%

Table 11 Accuracy for 4 cases for different number of epochs (15 frames)

The following graphs show the accuracy, for 15 frames, different epochs and cases.

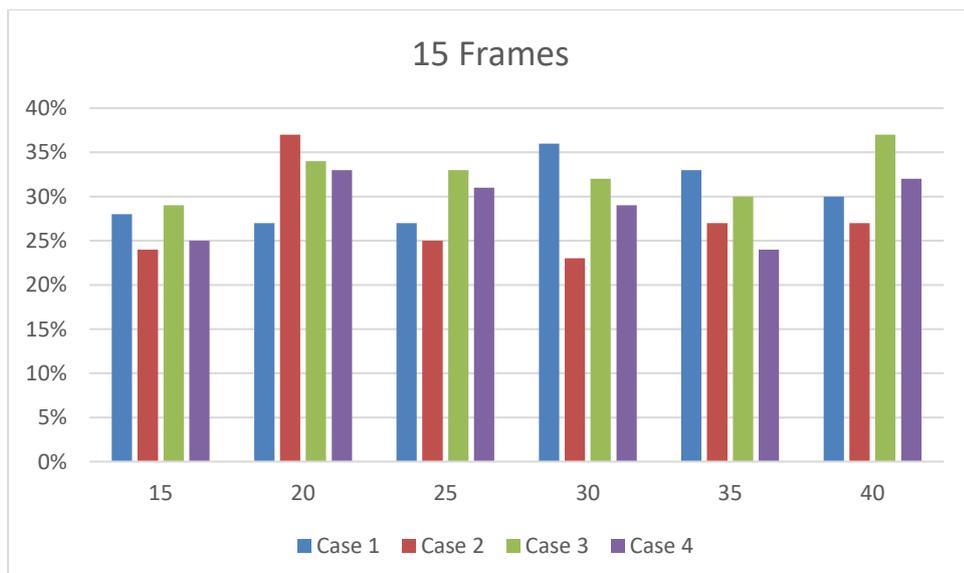
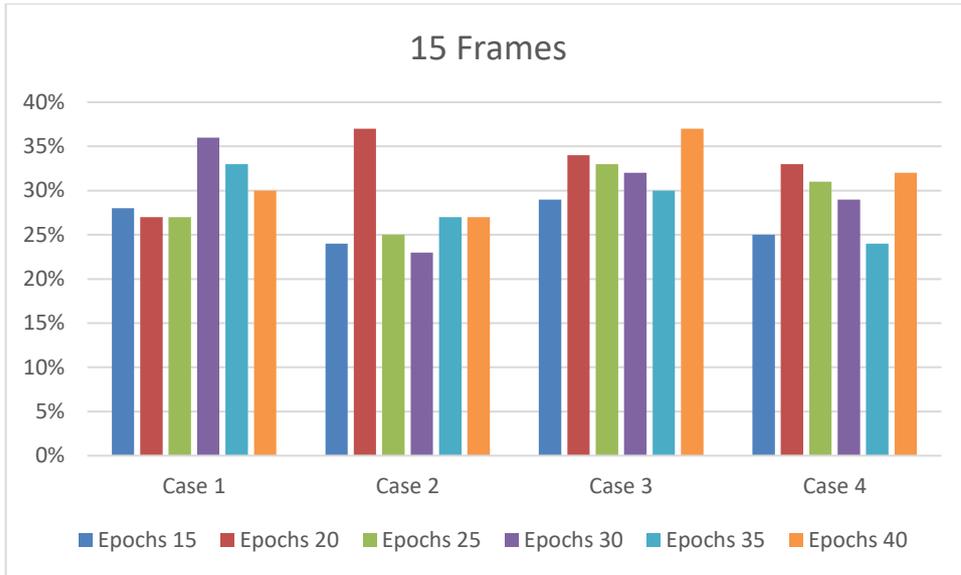


Figure 15 Case-wise accuracy comparison for different epochs (15 frames)



*Figure 16 Epoch wise accuracy comparison for different cases (15 frames)*

#### 4.2.2.2 For 20 frames

##### 1. Case 1 for LR 0.0001 and VR 0.2

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	29%	26%	27%	62%	59%	60%	23%	56%	33%	67%	7%	13%	24%	15%	18%
20	31%	41%	35%	34%	56%	42%	29%	22%	25%	45%	19%	26%	42%	37%	39%
25	32%	59%	42%	55%	63%	59%	37%	26%	30%	50%	4%	7%	36%	44%	40%
30	32%	33%	33%	35%	41%	38%	32%	22%	26%	25%	7%	11%	35%	63%	45%
35	39%	41%	40%	34%	78%	47%	40%	22%	29%	25%	4%	6%	42%	41%	42%
40	53%	30%	38%	47%	74%	57%	30%	37%	33%	11%	11%	11%	53%	33%	41%

Table 12 Precision, Recall, and F1 score for Case 1(20 frames)

For 20 frames, for 0.001 learning rate and 0.2 validation ratio, Dental class shows very low performance for all three parameters with fluctuations for all epochs. At 20 epochs, accuracy seems consistent across all the classes. The performance is swinging across the classes as the number of epochs increases. The following graph shows the summarized performance.

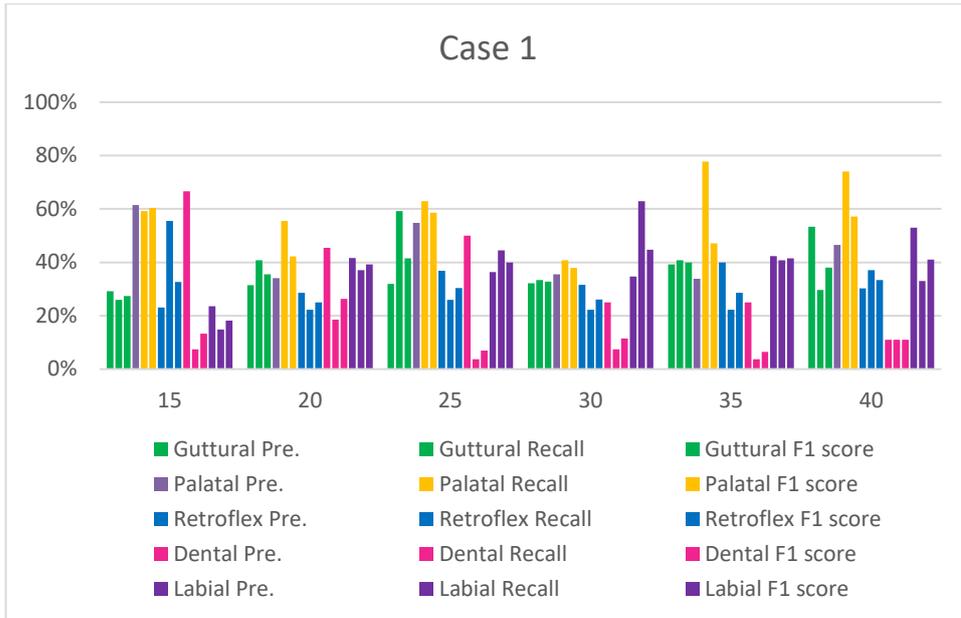


Figure 17 Accuracy comparison for different classes for case 1 (20 Frames)

2. Case 2 for LR 0.0002 and VR 0.2

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	25%	56%	34%	50%	4%	7%	21%	52%	29%	50%	4%	7%	67%	7%	13%
20	33%	4%	7%	32%	44%	38%	23%	26%	25%	22%	48%	30%	83%	19%	30%
25	50%	26%	34%	44%	59%	51%	33%	19%	24%	23%	11%	15%	33%	70%	45%
30	30%	33%	32%	57%	15%	24%	25%	48%	33%	11%	4%	6%	36%	48%	41%
35	33%	7%	12%	42%	37%	39%	30%	33%	32%	22%	41%	29%	35%	33%	34%
40	44%	26%	33%	35%	70%	46%	32%	26%	29%	44%	15%	22%	45%	56%	50%

Table 13 Precision, Recall, and F1 score for Case 2(20 frames)

Here, the performance parameters for retroflex class are not fluctuating for different epochs. However, from 15 to 35 epochs, except Retroflex, it is highly varying. The parameters values are very good for Labial and Palatal classes for 40 epochs. The following graph shows the performance of 20 frames with learning rate 0.002 and validation ratio 0.2.

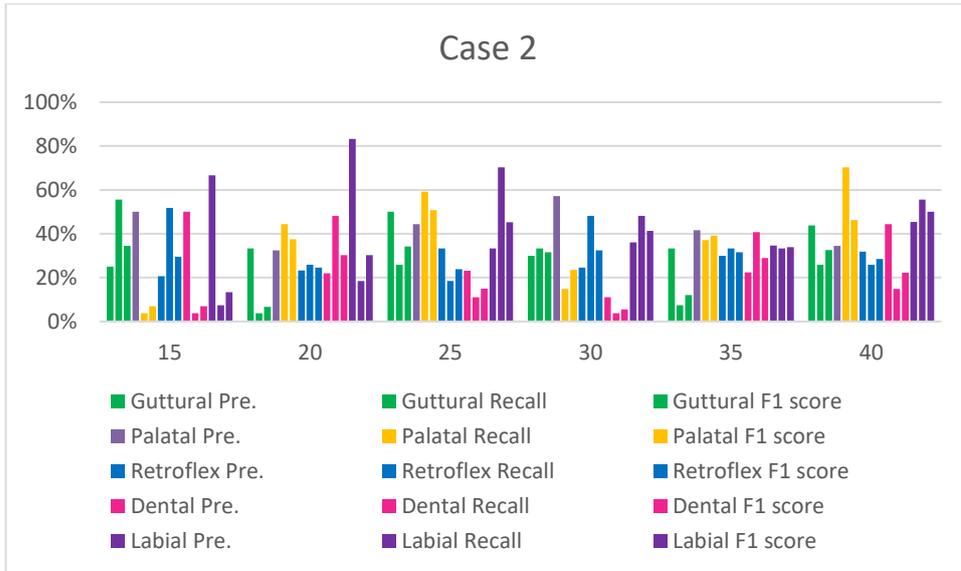


Figure 18 Accuracy comparison for different classes for case 2 (20 Frames)

### 3. Case 3 for LR 0.0001 and VR 0.25

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	31%	44%	36%	48%	59%	53%	33%	7%	12%	14%	15%	14%	32%	33%	33%
20	27%	44%	33%	36%	59%	44%	33%	7%	12%	11%	4%	6%	33%	37%	35%
25	28%	70%	40%	56%	56%	56%	27%	15%	19%	50%	19%	27%	31%	19%	23%
30	46%	22%	30%	50%	59%	54%	30%	52%	38%	25%	15%	19%	43%	44%	44%
35	38%	44%	41%	50%	63%	56%	47%	26%	33%	19%	19%	19%	43%	44%	44%
40	38%	52%	44%	60%	44%	51%	31%	37%	34%	18%	22%	20%	42%	19%	26%

*Table 14 Precision, Recall, and F1 score for Case 3(20 frames)*

The precision, recall and F1 score are giving fluctuating performance for 15 and 20 epochs for Retroflex and Dental classes. Performance is steady and balanced for 25 epochs onwards. The following graph shows the performance for case 3.

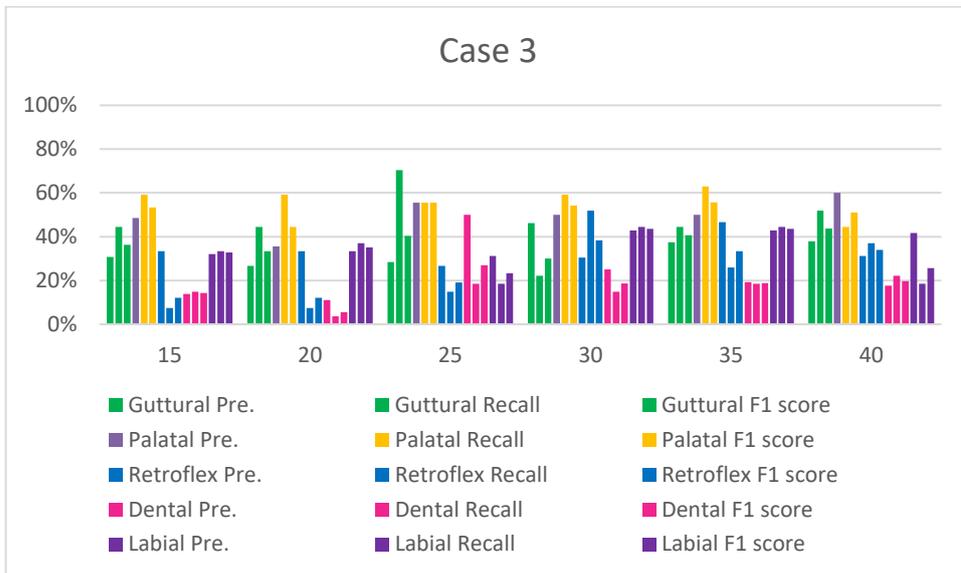


Figure 19 Accuracy comparison for different classes for case 3 (20 Frames)

4. Case 4 for LR 0.0002 and VR 0.25

Epochs	Guttural			Palatal			Retroflex			Dental			Labial		
	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score	Pre.	Recall	F1 score
15	38%	19%	25%	35%	78%	48%	20%	7%	11%	75%	11%	19%	29%	52%	37%
20	28%	56%	37%	33%	59%	43%	31%	15%	20%	10%	4%	5%	40%	15%	22%
25	23%	11%	15%	35%	70%	47%	100%	4%	7%	26%	33%	29%	28%	33%	31%
30	33%	48%	39%	64%	26%	37%	28%	37%	32%	23%	30%	26%	23%	11%	15%
35	50%	19%	27%	39%	41%	40%	29%	48%	36%	17%	22%	19%	44%	26%	33%
40	60%	11%	19%	43%	48%	46%	23%	26%	24%	8%	4%	5%	28%	59%	38%

Table 15 Precision, Recall, and F1 score for Case 4(20 frames)

The performance parameters for case 4 fluctuate highly for 15, 20, 35 and 40 epochs for Guttural, Retroflex, Dental classes. Performance parameters for Palatal class are balanced for all epochs.

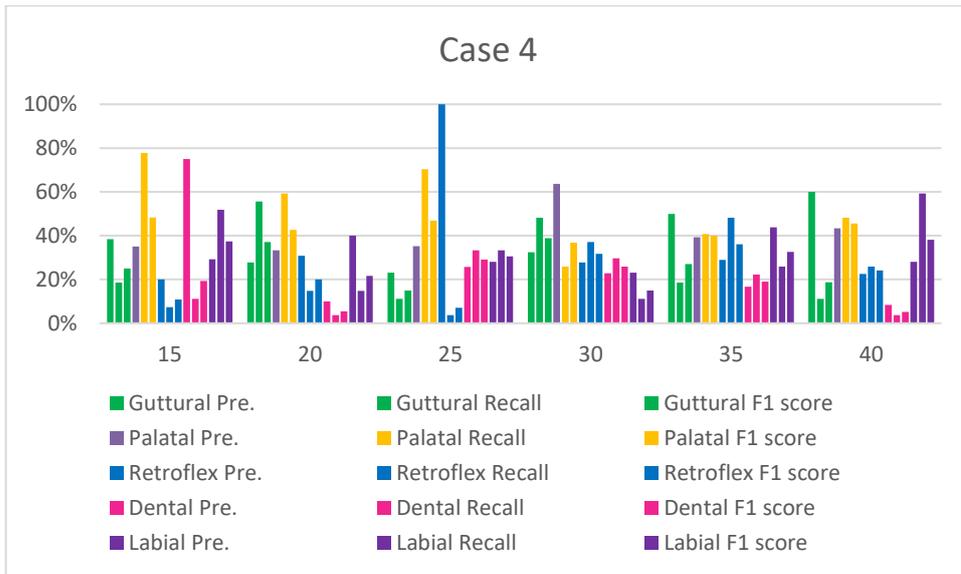


Figure 20 Accuracy comparison for different classes for case 4 (20 Frames)

The following graphs show the accuracy, for 20 frames, different epochs and cases.

Epochs	Case 1	Case 2	Case 3	Case 4
15	33%	24%	32%	33%
20	35%	28%	30%	30%
25	39%	37%	36%	30%
30	33%	30%	39%	30%
35	37%	30%	39%	31%
40	37%	39%	35%	30%

Table 16 Accuracy for 4 cases for different number of epochs (20 frames)

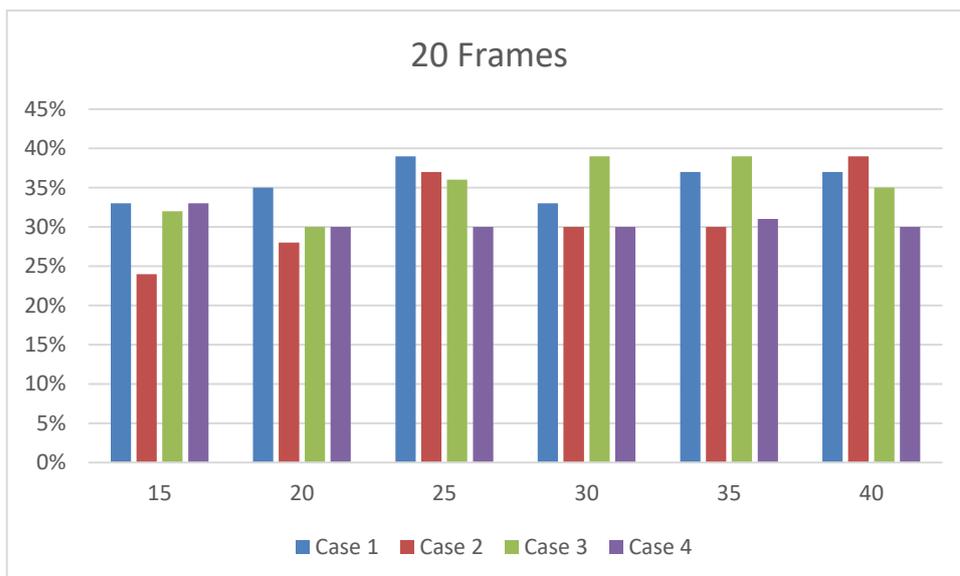
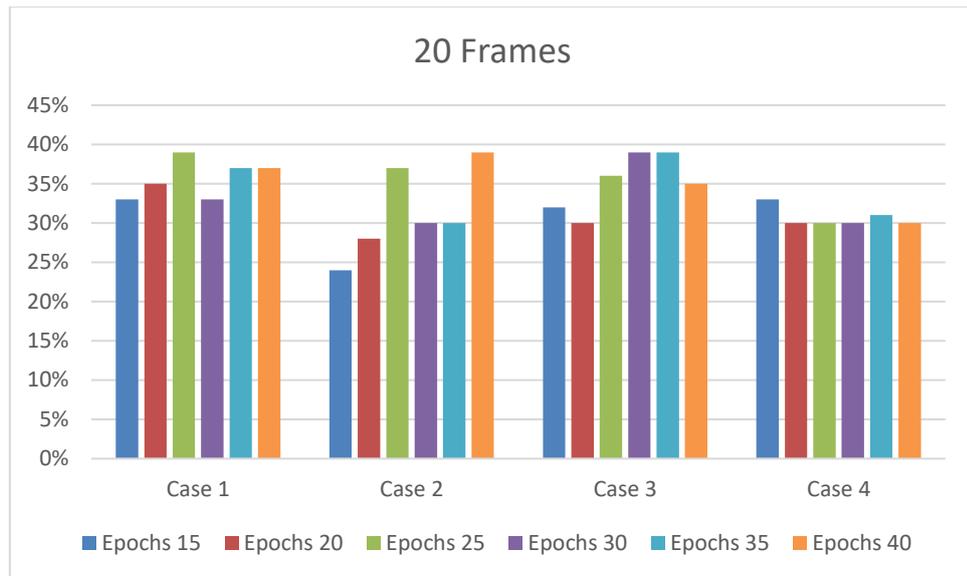


Figure 21 Case wise accuracy comparison for different epochs (20 frames)



*Figure 22 Epoch wise accuracy comparison for different cases (20 frames)*

## SUMMARY

This chapter carried out experimental analysis of the proposed ViLiDEx algorithm. The accuracy is good for 20 epochs, in case 1 for frame size 15 and 20. For case 2, accuracy is good for epochs 40 and for both frame sizes. For case 3, accuracy is good from 30 epochs onwards for 15 frames and 25 onwards for 20 frames. For case 4, accuracy is balanced for 35 and 40 epochs for 15 frames and the same is balanced for 30 and 35 epochs for 20 frames. Overall accuracy is good for validation ratio 0.25 for epochs 30 and 35.

