## **CHAPTER 5**

## **EXPERIMENTAL RESULTS**

In order to determine the accuracy of the system, five paragraphs containing Kannada text with distinct fonts are chosen and each paragraph output is predicted. The mis predicted words are highlighted. Errors occur either due to imperfect segmentation or due to misprediction by the CNN.

Imperfect segmentation results in character images which consist of more than one Akshara. Two Main Aksharas, two Vatt Aksharas or a Main Akshara and a Vatt Akshara may be present in one image. Also, an Akshara may not be properly segmented and may be split between two images. All these scenarios lead to a situation where the CNN definitely mis predicts and this leads to errors. Another cause of errors is pure misprediction by the CNN wherein the segmentation is perfect and a clean character image is given to the CNN for prediction, however the CNN mis predicts it as another character which is very similar to the original character.

Segmentation Accuracy is a measure of how well the characters are segmented and Prediction Accuracy is a measure of the overall accuracy of the system which includes the segmentation part and prediction by the CNN. Segmentation Accuracy is defined as the ratio of perfectly segmented characters to the total number of characters present in the test paragraph. Prediction Accuracy is defined as the ratio of correctly predicted words to the total number of words present in the paragraph. A word is considered to be erroneous even if a single character in it is mis predicted.

# **5.1 Sample and Output**

## 5.1.1 Paragraph 1

The sample image which was given as the input for the system is shown below

ನೂರಾರು ಪ್ರಾಣಿ ಪಕ್ಷಿಗಳ ಹೆಸರುಗಳು ದಟ್ಟೈಸಿವೆ. ಮರಗಿಡಗಳ ಪ್ರಸ್ತಾಪವಾಗುತ್ತದೆ. ಮೇಲಿನವುಗಳಲ್ಲದೆ ಅಸಂಖ್ಯಾತ ಅನಾಮಿಕ ಪಾತ್ರಗಳು, ಜಾಗಗಳು ಪ್ರಸ್ತಾಪವಾಗುತ್ತವೆ. ಇಷ್ಟೊಂದು ಪಾತ್ರಗಳು ಅವುಗಳ ಮನೋಭಾವ, ಕಾರ್ಯವಿಧಾನ, ವೃತ್ತಿ, ಸಂಬಂಧಗಳು ಎಲ್ಲವನ್ನೂ ಕವಿ ನಿಭಾಯಿಸಿರುವುದು ಅದ್ಭುತ!

ಎಲ್ಲಿಯಾದರೂ ಸರಿ, ಹೇಗಾದರೂ ಸರಿ ಒಂದಿಷ್ಟೂ ಗೊಂದಲ ಕಾದಂಬರಿಕಾರನಿಗೆ ಬಂದಿರಬಹುದಲ್ಲ ಎನ್ನುವ ಸಹಜ(ಕೆಟ್ಟ)ಕುತೂಹಲ ಉತ್ತರವಾಗಿ ಸಿಕ್ಕಿದ್ದು ಎರಡು ಸನ್ನಿವೇಶಗಳು. ಅದರಲ್ಲು ಒಂದು ಸನ್ನಿವೇಶ ಕಾದಂಬರಿಕಾರನ ಉದ್ದೇಶಪೂರ್ವಕ ನಡೆ ಎಂದು ಸ್ಪಷ್ಟವಾಗಿ ಗೊತ್ತಾಗುತ್ತದೆ. ಮತ್ತೊಂದು ಇಬ್ಬರು ವ್ಯಕ್ತಿಗಳ ಹೆಸರು ಮತ್ತು ವೃತ್ತಿಯ ಅದಲು ಬದಲು ಅಷ್ಟೆ!

ಈ ಕೆಳಗೆ ನಾನು ಪಟ್ಟಿ ಮಾಡಿರುವ ಸ್ಥಳಗಳು, ಪಾತ್ರಗಳು, ಅವುಗಳ ಪರಸ್ಪರ ಸಂಬಂಧಗಳನ್ನು ಗಮನಿಸಿದರೆ ಮಹಾಕಾದಂಬರಿಯೊಂದರ ಅರಹು ಹೇಗಿರುತ್ತದೆ ಹಾಗೂ ಹೇಗಿರಬೇಕು ಎಂದು ತಿಳಿಯುತ್ತದೆ. ಮಲೆಗಳಲ್ಲಿ ಮದುಮಗಳು ಓದುವುದಕ್ಕೆ ಪೂರ್ವಭಾವಿಯಾಗಿ ಈ ಸಿದ್ದಟಿಪ್ಪಣಿ ಒಳ್ಳೆಯ ಪ್ರವೇಶವಾಗಬಹುದು.

The Editable Output Text is as shown below. The errors have been highlighted manually

ಪ್ರಾಣಿ ಪಕ್ರಿಗಳ ಹೆಸರುಗಳು ದಟ್ಟಿಸಿವೆ ಮರಗಿಡಗಳ ನೂರಾರು ಪ್ರಸ್ತಾಪವಾಗುತ್ತದೆ ಮೇಲಿನವುಗಳಲ್ಲದೆ ಅಸಂಖ್ಯಾತ ಅನಾಮಿಕ ಪಾತ್ರಗಳು ಜಾಗಗಳು ಪ್ರಸ್ತಾಪವಾಗುತ್ತವೆ ಇಷ್ಟೊಂದು ಪಾತ್ರಗಳು ಅವುಗಳ ಮನೋಬಾವ <mark>ಕಾರ್ಯವಿದಾನ</mark> ವೃತ್ತಿ <mark>ಸಊಂದಗಳು</mark> ಎಲ್ಲವನ್ನೂ ಕವಿ ನಿ<mark>ಬಾಯಿಸಿರುವುದು</mark> <mark>ಅದ್ಭುತ್</mark>ಕ ಎಲ್ಲಿಯಾದರೂ ಸರಿ ಹೇಗಾದರೂ ಸರಿ ಒಂದಿಷ್ಟೂ ಗೊಂದಲ ಕಾದಂಬರಿಕಾರನಿಗೆ ಬಂದಿರಬಹುದಲ್ಲ ಸಹಜ(ಕೆಟ್ಟ)ಕುತೂಹಲ ಎನ್ನುವ ಉತ್ತರವಾಗಿ ಸಿಕ್ಕಿದ್ದು ಎರಡು ಸನ್ನಿವೇಶಗಳು ಅದರಲ್ಲು ಒಂದು ಸನ್ನಿವೇಶ ಕಾದಂಬರಿಕಾರನ ಉದ್ದೇಶಪೂರ್ವಕ ನಡೆ ಎಂದು ಸೃಷ್ಟವಾಗಿ ಗೊತ್ತಾಗುತ್ತದೆ ಮತ್ತೊಂದು ಇಬ್ಬರು ವ್ಯಕ್ತಿಗಳ ಹೆಸರು ಮತ್ತು ವೃತ್ತಿಯ ಅದಲು ಬದಲು <mark>ಅರ್ಷ್ಟ</mark>ೇ ಈ ಕೆಳಗೆ ನಾನು ಪಟ್ಟಿ ಮಾಡಿರುವ ಸ್ಥಳಗಳು ಪಾತ್ರಗಳು ಅವುಗಳ ಪರಸ್ಪರ ಸಂಬಂದಗಳನ್ನು ಗಮನಿಸಿದರೆ ಮಹಾಕಾದಂಬರಿಯೊಂದರ ಅರಹು ಹೇಗಿರುತ್ತದೆ ಹಾಗೂ ಹೇಗಿರಬೇಕು ಎಂದು ತಿಳಿಯುತ್ತದೆ ಮಲೆಗಳಲ್ಲಿ ಮದುಮಗಳು ಓದುವುದಕ್ಕೆ <mark>ಪೂರ್ವಬಾವಿಯಾಗಿ</mark> ಈ ಸಿದ್ಯಟಿಪ್ಪಣಿ ಒಳ್ಳೆಯ ಪ್ರವೇಶವಾಗಬಹುದು

#### The Editable Unicode output of the sample is as shown below:

\u0CA80CC2\u0CB00CBE\u0CB00CC1 \u0CAA0CCD0CB00CBE\u0CA30CBF \u0CAA\u0C950CCD0CB00CBF\u0C97\u0CB3 \u0CB00CC1\u0C97\u0CB3\u0C1\u0C97\u0CB3\u0C1\u0C97\u0CB3\u0C1\u0C97\u0CB3\u0C1\u0C97\u0CB3\u0C1\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0CB3\u0C97\u0C97\u0CB3\u0C97\u0C97\u0CB3\u0C97 \u0CA6\u0C9F0CCD0CB00CC6\u0CB80CBF\u0CB50CC6 \u0CAE\u0CB0\u0C970CBF\u0CA1\u0C97\u0CB3 \u0CAA0CCD0CB0\u0CB80CCD0CA40CBE\u0CAA\u0CB50CBE \u0C970CC1\u0CA40CCD0CA4\u0CA60CC6 \u0CAE0CC6\u0CD5\u0CB20CBF\u0CA8\u0CB50CC1\u0C97\u0CB3\u0CB20CCD0CB2\u0CA60CC6 \u0C85\u0CB8\u0C82\u0C960CCD0CAF0CBE\u0CA4 \u0C970CC1\u0CA40CCD0CA4\u0CB50CC6 \u0C87\u0CB70CCD0C9F0CCA\u0C82\u0CA60CC1 \u0CA40CBE\u0CA40CCD0CB0\u0C97\u0CB30CC1 \u0C85\u0CB50CC1\u0C97\u0CB3 \u0CAE \u00CA80CCA\u0CD5\u0CAC0CBE\u0CB5\u0CB5\u0CBE\u0CB5\u0CB6\u0CB5\u0C \u0C8E\u0CB20CCD0CB2\u0CB5\u0CA80CCD\u0CA80CC2\u0CB5\u0CB50CBF\u0CA80CC1\u0CB50CBF\u0CA60CBF\u0CA60CBF\u0CA60CBF\u0CA60CBF\u0CA60CBF\u0CB50CBF\u0C \u00085\u00C460CCD0CAD0CC1\u00C4\u00CEF\u00CB20CCD0CB20CBF\u00CA6\u00CA0\u00CA0\u00CB00CC2 \u00CB8\u00CB00CBF \u00CB90CC6\u00CD5\u00C970CBE\u00CB0\u00CB00CC2 \u0CAC\u0C82\u0CA6\u6C8\u0CB0\u0CB0\u0CAC\u0CB90CC1\u0CA6\u0CB20CCD0CB2 \u0C8E\u0CA80CCD0CA80CC1\u0CB5 \u0CB8\u0CB9\u0C9C\u28.0\u0C950CC6\u0C9F0CCD0C9F \u29.0\u000050CC1\u00CA40CC2\u00E9\u00E9\u00E0 \u00E8\u00E9\u00E04\u00E00\u00E00\u00E50CBE\u00E970CBF \u00E80CBF\u00E950CCD0C950CBF\u00E40CCC1\u00E86CCCD0CA60CC1 \u00E8E\u00E80\u00EA10CC1 \u0CB8\u0CA80CCD0CA80CBF\u0CB50CC6\u0CD5\u0CB6\u0C97\u0CB30CC1 \u0CB3\u0CA80CD0CA80CBF \u0CB2\u0CA80CD0CA80CBT \u0CB8\u0CA80CCD0CA80CBF \u0CA8\u0CA10CC6 \u0C8\u0C82\u0C82\u0CA60CC1 \u0CB80CCD0CAA\u0CB70CCD0C9F\u0CB50CBE\u0C970CCF \u0C970CCA\u0CA40CCD0CA40CCD0CA40CCD0CA4\u0CA60CC6 \u00CAE\u00CA40CCD0CA40CCA\u00C82\u00CA60CC1 \u00C87\u00CACOCCD0CAC\u00E800CC1 \u00CB50CCD0CAF\u00C950CCD0CAF\u00C950CDDCAA\u00C97\u00C93 \u00CB90CC6\u00C88\u00CB00CC1 \u00CAE \u0CA40CCD0CA40CC1 \u0CB50CC30.0\u0CA40CCD0CA40CEF\u0CAF\u0CAF\u0CAF\u0CAF\u0CAF\u0CB20CC1 \u0CAC\u0CA6\u0CB20CC1 \u0CAC\u0CB5\u0CB70CCD0C9F0CC6\u0CEF \u0C88 \u0CAA0CBE\u0CA40CCD0CB0\u0C97\u0CB30CC1 \u0C85\u0CB50CC1\u0C97\u0CB3 \u0CAA\u0CB0\u0CB80CCD0CAA\u0CB0 \u0CB8\u0CB2\u0CAC \u0082\u00C86\u00C97\u00E8\u00C880CCD0CA80CC1 \u00C97\u00E8\u00e8\u \u0C82\u0CA6\u0CB0 \u0C85\u0CB0\u0CB9\u0CB9\u0C6\u0CD5\u0C970CBF\u0C970CBF\u0CB0\u0CD5\u0C970CBF\u0CB0\u0CB0\u0CB9\u0CB0\u0CB0\u0CB9\u0CB0\u0CB9\u0CB0\u0CB9\u0CB0\u0CB9\u0CB0\u0CB9\u0CB0\u0CB9 \u0CB0\u0CA60CC6\u0CD5\u0C950CC1 \u0C8E\u0C82\u0CA60CC1 \u0CA40CBF\u0CB30CBF\u0CAF0CC1\u0CA40CCD0CA4\u0CA60CC6 \u0CAE\u0CB20CC6\u0C97\u0CB3\u0CB20CCD0CB20CBF \u0CAE\u0CAE\u0CAE\u0CAE\u0C97\u0CB50CC1 \u0C93\u0CA50CC1\u0CB50CCT\u0CB50CCB\u0C970CBF\u0CAF0CBE\u0C970CBF \u0C88 \u0C880EF\u0C460CCD0CA7\u0C9F0CBF\u0CA40CCD0CA4\u0CA30CBF \u0C92\u0CB30CC6\u0CAF \u0C92\u0CB30CC6\u0CAF \u0CA40CCD0CB0\u0CB50CE6\u0CB50CB6\u0CB50CBE\u0C97\u0CA \u0CB90CC1\u0CA60CC1

The paragraph consists of 82 words and 318 characters. 73 words have been predicted correctly and 300 characters have been segmented perfectly. Out of the eighteen imperfectly segmented characters, twelve are punctuations. Majority of the errors are caused due to imperfect segmentation. The Segmentation Accuracy is 94% and Prediction Accuracy is 89%.

# **Imperfect Segmentation**

Word	<b>Imperfect Character Segmentation</b>					
ದಟ್ಟೈಸಿವೆ	ದ img40_main	<b>战</b> img41_main	ಡ್ರ) img42_vatt	స్టి img43_main	ವೆ img44_main	n
ಕಾರ್ಯವಿಧಾನ,	<b>ප</b> img200_main	ಯ img201_main	E img202_main	ධ img203_main	ದಾ img204_main	え img205_main
ನಿಭಾಯಿಸಿರುವುದು	ට img250_main im	ಬಾ 1g251_main img	_	ಸಿ ರು :53_main img254_m	ವು ain img255_main	ದು img256_main



In the 1<sup>st</sup> example, both the Vatt Aksharas are saved as a single image and not further segmented. This is due to the fact that since they are in close proximity, they are treated as a single Connected Component. In the remaining examples, the "small lines" under Aksharas like "bha", "ra" and exclamation sign are not considered while segmentation.

# Misprediction by the CNN

Word	Original Akshara	Mispredicted Akshara		
ಪಕ್ಷಿಗಳ	ω,	<u>ن</u>		

## 5.1.2 Paragraph 2

The sample image which was given as the input for the system is shown below

ಕ್ಯಾನ್ಸರ್ ಇರುವುದು ಪತ್ತೆಯಾದಾಗ ವಿನಾಶಕಾರಿ ಭೀತಿ ಎದುರಾಗುತ್ತದೆ. ಆಘಾತ, ಯಾತನೆ, ಗೊಂದಲ, ಹತಾಶೆ... ಹೀಗೆ ಕ್ಯಾನ್ಸರ್ ಪತ್ತೆಯಾದಾಗ ಅದು ರೋಗಿಗಳು ಮತ್ತವರ ಪ್ರಿಯ ಜನರಿಂದ ಬದುಕನ್ನೇ ಕದ್ದೊಯ್ಡು ಬಿಡುತ್ತದೆ.

ಚಿಕಿತ್ಸೆ ಇನ್ನು ಕ್ಷಿಷ್ಟಕರವಾಗುವ ಕಾರಣವೇನೆಂದರೆ, ಬಹುಪಾಲು ರೋಗಿಗಳು, ರೋಗ ಬಲಿತ ನಂತರ ರೋಗವನ್ನು ಎದುರಿಸಲು ಮತ್ತು ನಿರ್ವಹಿಸಲು ಲಭ್ಯ ಚಿಕಿತ್ಸಾ ವಿಧಾನಗಳ ಮಾಹಿತಿಯ ಕೊರತೆಯಿಂದ ಬರುತ್ತಾರೆ.

The Editable Output Text is as shown below. The errors have been highlighted manually

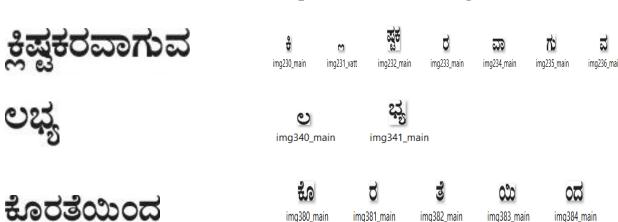
ಕ್ಯಾನ್ಸರ್ ಇರುವುದು ಪತ್ತೆಯಾದಾಗ ವಿನಾಶಕಾರಿ ಭೀತಿ ಎದುರಾಗುತ್ತದೆ ಆಫಾತ. ಯಾತನ. ಗೊಂದಲ. ಹತಾಶೆ ಹೀಗೆ ಕ್ಯಾನ್ಸರ್ ಪತ್ತೆಯಾದಾಗ ಅದು ರೋಗಿಗಳು ಮತ್ತವರ ಪ್ರಿಯ ಜನರಿಂದ ಬದುಕನ್ನೇ ಕದ್ದೊಯ್ಡು ಬಿಡುತ್ತದೆ ಚಿಕಿತ್ಸೆ ಇನ್ನು ಕ್ಲಿಕರವಾಗುವ ಕಾರಣವೇನೆಂದರೆ. ಬಹುಪಾಲು ರೋಗಿಗಳುಕೌರೋಗ ಬಲಿತ ನಂತರ ರೋಗವನ್ನು ಎದುರಿಸಲು ಮತ್ತು ನಿರ್ವಹಿಸಲು ಅ9 ಚಿಕಿತ್ಸಾ ವಿಧಾನಗಳ ಮಾಹಿತಿಯ ಕೊರತೆಯಿಯ ಬರುತ್ತಾರೆ

The paragraph consists of 40 words and 165 characters. 34 words have been predicted correctly and 156 characters have been segmented perfectly. Out of the nine imperfectly segmented characters, six are punctuations. Errors are equally contributed due to imperfect segmentation and misprediction by the CNN. Segmentation Accuracy is 94% and Prediction Accuracy is 85%.

# **Imperfect Segmentation**

#### Word

# **Imperfect Character Segmentation**



In the 1<sup>st</sup> example, "img232\_main" consists of two Main Aksharas and a Vatt Akshara. In the 2<sup>nd</sup> example, "img341\_main" consists of both the Main Akshara and the corresponding Vatt Akshara. In the 3<sup>rd</sup> example, "img384\_main" consists of two Main Aksharas. In all the cases, segmentation is imperfect due to very close proximity between the aksharas.

## Misprediction by the CNN

Word	Original Akshara	Mispredicted Akshara		
ಯಾತನೆ,	ನೆ	ನ		
ರೋಗಿಗಳು,	9	ಕೌ		

In the  $1^{st}$  example, "ne" is mis predicted as "na" which is understandable since both the characters are very similar. In the  $2^{nd}$  example, "coma" is mis predicted as "kou" which is a serious error.

## 5.1.3 Paragraph 3

The sample image which was given as the input for the system is shown below

ಆದರೀಗ, ಕಳೆದ ಐದು ದಶಕಗಳಿಂದ ಡಿ.ಎಸ್. ರಿಸರ್ಚ್ ಸೆಂಟರ್ನ ತಜ್ಞರು ರೋಗವನ್ನು ಧೈರ್ಯದಿಂದ ಎದುರಿಸಿ, ನಂತರ ಉತ್ತಮ ಬದುಕು ಸಾಗಿಸಲು ಅನೇಕ ಕ್ಯಾನ್ಸರ್ ಪೀಡಿತರಿಗೆ ನೆರವಾಗುತ್ತಿರುವುದರಿಂದ, ರೋಗಿಗಳಿಗೆ ಪ್ರಾಚೀನ ಆಯುರ್ವೇದಾಧಾರಿತ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ವಿಶ್ವಾಸವಿಡಲು ಪ್ರಬಲ ಕಾರಣಗಳಿವೆ. ರೋಗದ ಸಕಾಲಿಕ ನಿರ್ವಹಣೆಯೊಂದಿಗೆ 4ನೇ ಹಂತದ ಅನೇಕ ರೋಗಿಗಳ ಜೀವತಾವಧಿ ವಿಸ್ತರಣೆ ಮತ್ತು ಸಾಮಾನ್ಯ ಆರೋಗ್ಯ ಸುಧಾರಣೆಗಳಿಗೂ ಈ ಚಿಕಿತ್ಸೆ ನೆರವಾಗಿದ್ದು,

The Editable Output Text is as shown below. The errors have been highlighted manually

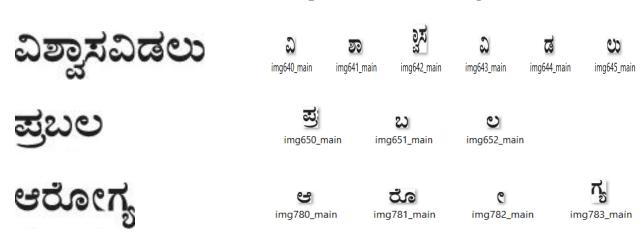
ಆದರೀಗ! ಕಳೆದ ಐದು ದಶಕಗಳಿಂದ ಡಿಎಸ್ ರೋಗವನ್ನು <mark>ಭೃತರ್ಯದಿಂದ</mark> ಎದುರಿಸಿ. ನಂತರ ಉತ್ತಮ ಬದುಕು ಸಾಗಿಸಲು ಪೀಡಿತರಿಗೆ ನೆರವಾಗುತ್ತಿರುವುದರಿಂದ. ಅನೇಕ ಕ್ಯಾನ್ಸರ್ ರೋಗಿಗಳಿಗೆ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ಆಯುವರ್ೕದಾದಾರಿತ ವಿಶಾಳವಿಡಲು ಕಾರಣಗಳಿವೆ ರೋಗದ ಸಕಾಲಿಕ ನಿರ್ವಹಣೆಯೊಂದಿಗೆ 4ನೇ ಹಂತದ ಅನೇಕ ರೋಗಿಗಳ ವಿಸ್ತರಣೆ ಮತು ಸಾಮಾನ್ಯ ದೇವತಾವದಿ ಸುಧಾರಣೆಗಳಿಗೂ ಈ ಚಿಕಿತ್ಸೆ ನೆರವಾಗಿದ್ದು

The paragraph consists of 42 words and 178 characters. 34 words have been predicted correctly and 171 characters have been segmented perfectly. Out of the seven imperfectly segmented characters, four are punctuations. Errors are equally contributed by imperfect segmentation and misprediction by the CNN. Further, some errors are due to flaw in printing. Segmentation Accuracy is 96% and Prediction Accuracy is 81%.

# **Imperfect Segmentation**

#### Word

# **Imperfect Character Segmentation**



In the 1<sup>st</sup> example, "img642\_main" contains part of the previous akshara including the Vatt Akshara and the current akshara. In the 2<sup>nd</sup> and 3<sup>rd</sup> example, "img650\_main" and "img783\_main" respectively consists of the Main Akshara and the Vatt Akshara. This is due to the close proximity of the aksharas and hence they are treated as a single Connected Component.

# Misprediction by the CNN

Word	Original Akshara	Mispredicted Akshara
ಸೆಂಟರ್ನ	ヹ	ぇ
ಜೀವತಾವಧಿ	ಜಿ	ದೆ

In the 1<sup>st</sup> example, "se" is mis predicted as "sa", which is understandable since both aksharas are very similar. However, in the 2<sup>nd</sup> example, "ji" is mis predicted as "de", which is a serious error.

## 5.1.4 Paragraph 4

The sample image which was given as the input for the system is shown below

ರೋಗಿಗಳಲ್ಲಿ ಯಕೃತ್ತು, ಮೆದುಳು, ಮೆದೋಜೀರಕ ಗ್ರಂಥಿ, ಮೆಟಾಸ್ಟಾಸಿಸ್ಗೆ ಕ್ರಮೇಣ ನಿರಾಶೆಗೆ ಬಲಿಯಾಗಲಿದ್ದವರೂ ಇದ್ದರು. ದೇಶದ ಒಂದಲ್ಲೊಂದು ಇವರುಗಳಿಗೆ 'ಭರವಸೆ ಕಳೆದ' ಆಸ್ಪತ್ರೆಯಲ್ಲಿ ರೋಗಿಗಳೆಂದು ನಿರಾಕರಿಸಲಾಗಿತ್ತೆಂಬುದು ಗಮನಿಸಬೇಕು. ಪ್ರಕರಣಗಳಿಂದು ಅಧ್ಯಯನ ಇದು ಕ್ಯಾನ್ಫರ್ ರೋಗಿಗಳಿಗೆ ಅಪಾರ ಶಕ್ತಿ ನೀಡುತ್ತದೆ, ಡಿಎನ್ಆರ್ಸಿಯಲ್ಲಿನ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ಭರವಸೆ ಹೊಂದಲು ಹಾಗೂ ನಿರ್ಧಾರದಿಂದ ಬದುಕಿ ಭಯ ದೂರ ಮಾಡಲು, ಕ್ಯಾನ್ನರ್ನೊಂದಿಗೆ ಸೆಣಸಿ, 'ಕ್ಯಾನ್ನರ್ ವಿಜೇತ'ರೆಂದೆನಿಸಲು ನೆರವಾಗುತ್ತದೆ.

The Editable Output Text is as shown below. The errors have been highlighted manually

ರೋಗಿಗಳಲ್ಲಿ ಮದುಳು. ಮೆದೋಜೀರಕ ඵංක. ಮೂತ್ರಪಿಂಡ. ಮೆಟಾಸ್ಕಾಸಿಸೆ. ಇರುವರಲ್ಲದೇ! ಒಳಗಾದವರೂ ಅವರು ಕಮೇಣ ನಿರಾಶೆಗೆ ಬಲಿಯಾಗಲಿದ್ದವರೂ ಇದ್ದರು ದೇಶದ ಒಂದಲ್ಲೊಂದು <mark>ಶಿನ್ಮರ್</mark> ಆಸತ್ರೆಯಲ್ಲಿ ಕಳೆದ. ರೋಗಿಗಳೆಂದು ನಿರಾಕರಿಸಲಾಗಿತ್ತೆಂಬುದು <mark>೬</mark>ಭರವಸೆ ಗಮನಿಸಬೇಕು ಪ್ರಕರಣಗಳಿಂದು ಅಧ್ಯಯನ ಈ ಯೋಢ್ವಾಗಿದ್ದು ಶಕ್ತಿ ನೀಡುತ್ತದೆ. ರೋಗಿಗಳಿಗೆ ಡಿಎನ್ಆರ್ಸಿಯಲ್ಲಿನ ಅಪಾರ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ಭರವಸೆ ಹೊಂದಲು ಹಾಗೂ ನಿರ್ಧಾರದಿಂದ ಬದುಕಿ ಭಯ ದೂರ ಮಾಡಲು, ಕ್ಯಾನ್ಫರ್ನೊಂದಿಗೆ ಸೆಣಸಿ. <mark>ಇ್ಯನ್ಫರ್</mark> ವಿಜೇತ,ರೆಂದೆನಿಸಲು ನೆರವಾಗುತ್ತದೆ

The paragraph consists of 52 words and 213 characters. 41 words have been predicted correctly and 207 characters have been segmented perfectly. Out of the six imperfectly segmented

characters, four are punctuations. Majority of the errors are caused due to imperfect segmentation. Segmentation Accuracy is 96% and Prediction Accuracy is 79%.

## **Imperfect Segmentation**

# Word

# **Imperfect Character Segmentation**



In examples one and two, "img871\_main" and "img881\_main" consist of part of the previous akshara along with the current akshara. Remaining examples consist of images with both the Main Akshara and the Vatt Akshara. Main Akshara and Vatt Akshara are considered as a single Connected Component due to their close proximity.

## Misprediction by the CNN

Word

# **Original Akshara**

**Mispredicted Akshara** 







"Open single inverted coma" is mispredicted as Kannada numeral "six". The CNN models have not been trained with "Open single inverted coma".

## 5.1.5 Paragraph 5

The sample image which was given as the input for the system is shown below

ನಾಡಿನ ಪುಣ್ಯದ ಪೂರ್ವ ದಿಗಂತದಿ ನವ ಅರುಣೋದಯ ಹೊಮ್ಮುತಿದೆ! ಚಿರನೂತನ ಚೇತನಾಡುತ್ಸಾಹದಿ ನವೀನ ಜೀವನ ಚಿಮ್ಮುತಿದೆ!

ಅಭಿನವ ಮಧುಕೋಕಿಲ ಕಲಾಕಂಠದಿ ಸ್ವರಸುರ ಚಾಪಗಳುಣ್ಮುತಿವೆ! ಶ್ಯಾಮಲ ಕಾನನ ಸುಮ ಸಮ್ಮೇಳದಿ ಇಂಚರ ಸಾಸಿರ ಪೊಣ್ಮುತಿವೆ! The Editable Output Text is as shown below. The errors have been highlighted manually

ನಾಡಿನ ಪುಣ್ಯದ ಪೂರ್ ದಿಗ0ತದಿ ನವ ಅರುಣೋದಯ ಹೊಮ್ಮುತಿದೆ! ಚೆರನೂತನ ಚೇತನಾಡುತ್ಸಾಹದಿ ನವೀನ ಜೀವನ ಚೆಮ್ಮುತಿದೆ ಅಭಿನವ ಮಧುಕೋಕಿಲ ಕಲಾಕ0ಠದಿ ಸ್ವರಸುರ ಚಾಪಗಳುಣ್ಮುತಿವೆ ಶ್ಯಾಮಲ ಕಾನನ ಸುಮ ಸಮೇನಳದಿ ಇಂಚರ ಸಾಸಿರ ಪೊಣ್ಣುತಿವೆ

The paragraph consists of 24 words and 115 characters. 18 words have been predicted correctly and 113 characters have been segmented perfectly. Both the imperfectly segmented characters are punctuations. All the errors are caused due to misprediction by the CNN. Segmentation Accuracy is 98% and Prediction Accuracy is 75%.

# Misprediction by the CNN

Word	Original Akshara	Mispredicted Akshara		
ದಿಗಂತದಿ	0	0		
	(Anuswara)	(English numeral zero)		
ಚಿರನೂತನ	ಚ	ಚೆ		
ಪೊಣ್ಮುತಿವೆ!	P	o.		

In the 1<sup>st</sup> example, Kannada Anuswara is mispredicted as English numeral "zero". In the 2<sup>nd</sup> example, akshara "chi" is mispredicted as "che". These are understandable since the corresponding

characters have lots of similarity. In the 3<sup>rd</sup> example, Vattu Akshara "ma" is mispredicted as "na" which is a serious error.

# **5.2 Model Accuracy**

Based on the Paragraph of Inputs and their predicted output as shown in previous section, the Prediction Accuracy and Segmentation Accuracy is as shown in the Table 5.1.

The accuracy is calculated using the following formula:

Prediction Accuracy (in %) = (Correct output/No of Input Words) \*100

Segmentation Accuracy (in %) = (Perfectly Segmented Characters/No of Input Characters) \*100

INPUT PARAGRAPH NO	NO. OF INPUT WORDS	CORRECT OUTPUT	NO. OF INPUT CHARACTERS	PERFECTLY SEGMENTED CHARACTERS	PREDICTION ACCURACY (IN %)	SEGMENTATION ACCURACY (IN %)
1	82	73	318	305	89	94
2	40	34	165	156	85	94
3	42	34	178	171	81	96
4	52	41	213	205	79	96
5	24	18	115	113	75	98
AVERAGE PREDICTION ACCURACY		82				
AVERAGE SEGMENTATION ACCURACY		96				

**TABLE 5.1 Model Accuracy Calculation**