
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 ದಟ್ಟಿಸಿವೆ. 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 ಸಿದ್ಧಿಪಟ್ಟಣಿ ಒಳ್ಳೆಯ ಪ್ರವೇಶವಾಗಬಹುದು.

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

ನೂರಾರು ಪ್ರಾಣಿ ಪಕ್ಷಿಗಳ ಹೆಸರುಗಳು ದಟ್ಟಿಸಿವೆ 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 ಸಿದ್ಧಿಪಟ್ಟಣಿ ಒಳ್ಳೆಯ ಪ್ರವೇಶವಾಗಬಹುದು

\\0C80CC2\\0CB00BE\\0CB00CC1 \\0CAABCCD0CB00CB\\0CA30CBF \\0CAA\\0C950CCD0CB00CB\\0C97\\0CB3 \\0CB90CC6\\0CB8\\0CB00CC1\\0C97\\0CB30CC1
\\0CA86\\0C9F0CCD0CB00CC\\0CB00CB\\0CB88CBF\\0CB50CC6 \\0CAE\\0CB0\\0C970CBF\\0CA1\\0C97\\0CB3 \\0CAA0CCD0CB0\\0CB80CCD0CA40BE\\0CAA\\0CB50CBE
\\0CB970CC1\\0CA40CCD0CA4\\0CA60CC6 \\0CAE0CC6\\0CD5\\0CB20CBF\\0CA8\\0CB50CC1\\0C97\\0CB3\\0CB20CCD0CB2\\0CA60CC6 \\0CB5\\0CB8\\0CB2\\0C960CCD0CAF0CBE\\0CA4
\\0CB5\\0CA80CBE\\0CAE0CBF\\0CB95 \\0CA0ABCE\\0CA40CCD0CB0\\0C97\\0CB30CC1 \\0C9C0CBE\\0C97\\0C97\\0CB30CC1 \\0CA0A0CCD0CB0\\0CB80CCD0CA40BE\\0CAA\\0CB50CBE
\\0CB970CC1\\0CA40CCD0CA4\\0CB50CC6 \\0CB87\\0CB870CCD0C9F0CCA\\0CB82\\0CA60CC1 \\0CAA0ABCE\\0CA40CCD0CB0\\0CB97\\0CB30CC1 \\0CB5\\0CB50CC1\\0C97\\0CB3 \\0CAE
\\0CB80CCA\\0CD5\\0CA0CBE\\0CB85 \\0C950CBE\\0CB80CCD0CB0CAF\\0CB50CBF\\0CA60CBE\\0CA8 \\0CB50CC30.0\\0CB440CCD0CA40BF \\0CB8\\0CB8\\0CB82\\0CA6\\0C97\\0CB30CC1
\\0CB85\\0CB20CCD0CB2\\0CB85\\0CA80CCD0CA80CC2 \\0C95\\0CB50CBF \\0CA80CBF\\0CA0CBE\\0CAFC0BE\\0CB88CBF\\0CB00CC1\\0CB550CC1\\0CA60CC1
\\0CB85\\0CA60CCD0AD0CC1\\0CA4\\0CEF \\0CB8E\\0CB20CCD0CB20CBF\\0CAFC0BE\\0CA6\\0CB00CC2 \\0CB8\\0CB80CBF \\0CB90CC6\\0CD5\\0CB970CBE\\0CA6\\0CB00CC2
\\0CB8\\0CB80CBF \\0C92\\0CB82\\0CA60CBF\\0CB870CCD0C9F0CC2 \\0C970CCA\\0CB82\\0CA6\\0CB2 \\0C950CBE\\0CA6\\0CB2\\0CB8\\0CB80CBF\\0CB950CBE\\0CB0\\0CA80CBF\\0CB970CC6
\\0CAC\\0CB2\\0CA60CBF\\0CB0\\0CAC\\0CB90CC1\\0CA6\\0CB20CCD0CB2 \\0CB8E\\0CA80CCD0CA80CC1\\0CB5 \\0CB8\\0CB9\\0C9C\\028.0\\0C950CC6\\0C9F0CCD0C9F
\\029.0\\0C950CC1\\0CA40CC2\\0CB9\\0CB2 \\0CB9\\0CA40CCD0CA4\\0CB0\\0CB50CBE\\0C970CBF \\0CB80CBF\\0C950CCD0C950CBF\\0CA60CCD0CA60CC1 \\0CB8E\\0CB0\\0CA10CC1
\\0CB8\\0CA80CCD0CA80CBF\\0CB50CC6\\0CD5\\0CB6\\0C97\\0CB30CC1 \\0CB5\\0CA6\\0CB8\\0CB20CCD0CB20CC1 \\0C92\\0CB2\\0CA60CC1 \\0CB8\\0CA80CCD0CA80CBF
\\0CB50CC6\\0CD5\\0CB6 \\0C950CBE\\0CA6\\0CB2\\0CAC\\0CB80CBF\\0C950CBE\\0CB0\\0CA8 \\0CB9\\0CA60CCD0CA60CC6\\0CD5\\0CB6\\0CA0ABCE\\0CB80CCD0CB0\\0CB95
\\0CA8\\0CA10CC6 \\0CB8E\\0CB2\\0CA60CC1 \\0CB80CCD0CAA\\0CB70CCD0C9F\\0CB50CBE\\0C970CBF \\0C970CCA\\0CA40CCD0CA40CBF\\0C970CC1\\0CA40CCD0CA4\\0CA60CC6
\\0CAE\\0CA40CCD0CA40CCA\\0CB2\\0CA60CC1 \\0CB87\\0CAC0CCD0CA\\0CB00CC1 \\0CB50CCD0CAF\\0C950CCD0CA40BF\\0C97\\0CB3 \\0CB90CC6\\0CB8\\0CB00CC1 \\0CAE
\\0CA40CCD0CA0CC1 \\0CB50CC30.0\\0CA40CCD0CA40CBF\\0CAF \\0CB85\\0CA6\\0CB20CC1 \\0CAC\\0CA6\\0CB20CC1 \\0CB5\\0CB70CCD0C9F0CC6\\0CEF \\0CB8
\\0C950CC6\\0CB4\\0C970CB \\0CA80CBE\\0CA80CC1 \\0CAF\\0C9F0CCD0C9F0CBF \\0CAE0CBE\\0CA10CBF\\0CB00CC1\\0CB5 \\0CB80CCD0CA7\\0CB3\\0C97\\0CB30CC1
\\0CA0ABCE\\0CA40CCD0CB0\\0C97\\0CB30CC1 \\0CB5\\0CB50CC1\\0C97\\0CB3 \\0CAA\\0CB0\\0CB80CCD0CAA\\0CB0 \\0CB8\\0CB82\\0CAC
\\0CB2\\0CA6\\0C97\\0CB3\\0CB80CCD0CA80CC1 \\0C97\\0CAE\\0CA80CBF\\0CB80CBF\\0CA6\\0CB00CC1 \\0CAE\\0CB90CBE\\0C950CBE\\0CB2\\0CB2\\0CAC\\0CB80CBF\\0CAFA0CCA
\\0CB2\\0CA6\\0CB0 \\0CB5\\0CB8\\0CB90CC1 \\0CB80CB6\\0CD5\\0C970CBF\\0CB00CC1\\0CA40CCD0CA4\\0CA60CC6 \\0CB90CBE\\0C970CC2 \\0CB90CC6\\0CD5\\0CB970CBF
\\0CB0\\0CAC0CC6\\0CD5\\0C950CC1 \\0CB8E\\0CB2\\0CA60CC1 \\0CA40CBF\\0CB30CBF\\0CAFA0CC1\\0CA40CCD0CA4\\0CA60CC6 \\0CAE\\0CB20CC6\\0C97\\0CB3\\0CB20CCD0CB20CBF
\\0CAE\\0CA60CC1\\0CAE\\0C97\\0CB30CC1 \\0CB93\\0CA60CC1\\0CB50CC1\\0CA6\\0C950CCD0C950CC6 \\0CA0CC2\\0CB80CCD0CB5\\0CAC0CBE\\0CB50CBF\\0CAFC0BE\\0C970CBF
\\0CB8 \\0CB80CBF\\0CA60CCD0CA7\\0C9F0CBF\\0CAA0CCD0CAA\\0CA30CBF \\0C92\\0CB30CCD0CB30CC6\\0CAF \\0CAA0CCD0CB0\\0CB50CC6\\0CD5\\0CB6\\0CB50CBE\\0C97\\0CAC
\\0CB90CC1\\0CA60CC1

Imperfect Segmentation

Imperfect Character Segmentation

ದ ಟ ಛ ಸಿ ವೆ
img40_main img41_main img42_vatt img43_main img44_main

ಕಾ	ಯ	ಫ	ವಿ	ದಾ	ನ
img200_main	img201_main	img202_main	img203_main	img204_main	img205_main

ನಿ ಬಾ ಮಿ ಸಿ ರು ವು ದು
img250 main img251 main img252 main img253 main img254 main img255 main img256 main

ಅದ್ಭುತ!

ಅ
img260_main

ದು
img261_main

ಛ
img262_vatt

ತ
img263_main

ಲ
img264_main

ಅಷ್ಟೆ!

ಅ
img590_main

ಷೆ
img591_main

ಛ
img592_vatt

ಲ
img593_main

ಪೂರ್ವಭಾವಿಯಾಗಿ

ಪೂ
img820_main

ವ
img821_main

ಫ
img822_main

ಬಾ
img823_main

ವಿ
img824_main

ಯಾ
img825_main

ಗಿ
img826_main

In the 1st 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		

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 ಕದ್ದೊಯ್ದು ಬಿಡುತ್ತದೆ.

ಚಿಕಿತ್ಸೆ ಇನ್ನು ಕ್ಲಿಷ್ಟಕರವಾಗುವ 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
ಆಘಾತ. ಯಾತನೆ. ಗೊಂದಲ. ಹತಾಶೆ ಹೀಗೆ ಕ್ಯಾನ್ಸರ್ ಪತ್ತೆಯಾದಾಗ ಅದು
ರೋಗಿಗಳು ಮತ್ತವರ ಪ್ರಿಯ ಜನರಿಂದ ಬದುಕನ್ನೇ ಕದ್ದೊಯ್ದು ಬಿಡುತ್ತದೆ
ಚಿಕಿತ್ಸೆ ಇನ್ನು ಕ್ಲಿಷ್ಟಕರವಾಗುವ ಕಾರಣವೇನೆಂದರೆ. ಬಹುಪಾಲು ರೋಗಿಗಳುಕೂ
ರೋಗ ಬಲಿತ ನಂತರ ರೋಗವನ್ನು ಎದುರಿಸಲು ಮತ್ತು ನಿರ್ವಹಿಸಲು ಲ9
ಚಿಕಿತ್ಸಾ ವಿಧಾನಗಳ ಮಾಹಿತಿಯ ಕೊರತೆಯಿಂದ 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

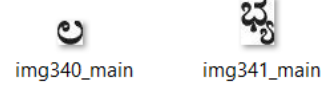
Word

ಕಿಷ್ಕಕರವಾಗುವ

ಲಭ್ಯ

ಕೊರತೆಯಿಂದ

Imperfect Character Segmentation



In the 1st example, “img232_main” consists of two Main Aksharas and a Vatt Akshara. In the 2nd example, “img341_main” consists of both the Main Akshara and the corresponding Vatt Akshara. In the 3rd 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,

ನಾ

ನಾ

ರೋಗಿಗಳು,

ಕೊ

ಕೊ

In the 1st example, “ne” is mis predicted as “na” which is understandable since both the characters are very similar. In the 2nd 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 ರಿಸರ್ ಸೆಂಟರ್ನ ತಜ್ಞರು ರೋಗವನ್ನು ಭೈರ್ಯದಿಂದ ಎದುರಿಸಿ. ನಂತರ ಉತ್ತಮ ಬದುಕು ಸಾಗಿಸಲು ಅನೇಕ ಕ್ಯಾನ್ಸರ್ ಪೀಡಿತರಿಗೆ ನೆರವಾಗುತ್ತಿರುವುದರಿಂದ. ರೋಗಿಗಳಿಗೆ ಪ್ರಾಚೀನ ಆಯುರ್ವೇದಾಧಾರಿತ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ವಿಶ್ವಾಸವಿಡಲು ತಬಲ ಕಾರಣಗಳಿವೆ ರೋಗದ ಸಕಾಲಿಕ ನಿರ್ವಹಣೆಯೊಂದಿಗೆ 4ನೇ ಹಂತದ ಅನೇಕ ರೋಗಿಗಳ ದೇವತಾವಧಿ ವಿಸ್ತರಣೆ ಮತ್ತು ಸಾಮಾನ್ಯ ಆರೋಗ್ಯ 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

ವಿಶ್ವಾಸವಿಡಲು

ಪ್ರಬಲ

ಆರೋಗ್ಯ

Imperfect Character Segmentation

ವಿ ಶಾ ವಿಸ್ ವಿ ಡ ಲು
img640_main img641_main img642_main img643_main img644_main img645_main

ಪ್ರ ಬ ಲ
img650_main img651_main img652_main

ಆ ರೋ ಗ್ಯ
img780_main img781_main img782_main img783_main

In the 1st example, “img642_main” contains part of the previous akshara including the Vatt Akshara and the current akshara. In the 2nd and 3rd 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

ಸೆಂಟರ್‌ನ

ಜೀವತಾವಧಿ

Original Akshara

ಸೆ

ಜಿ

Mispredicted Akshara

ಸಾ

ದಿ

In the 1st example, “se” is mis predicted as “sa”, which is understandable since both aksharas are very similar. However, in the 2nd 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

ಈ ರೋಗಿಗಳಲ್ಲಿ ಯಕೃತ್ತು, ಮೆದುಳು, ಶ್ವಾಸಕೋಶ, ಪ್ರಾಸ್ಟೇಟ್, ಮೆದೋಜೀರಕ ಗ್ರಂಥಿ, ಮೂತ್ರಪಿಂಡ, ಸೆರ್ವಿಕ್ಸ್‌ಗಳ ಮೆಟಾಸ್ಟಾಸಿಸ್‌ಗೆ ಒಳಗಾದವರೂ ಇರುವರಲ್ಲದೇ, ಅವರು ಕ್ರಮೇಣ ನಿರಾಶೆಗೆ ಪೂರ್ತಿ ಬಲಿಯಾಗಲಿದ್ದವರೂ ಇದ್ದರು. ದೇಶದ ಒಂದಲ್ಲೊಂದು ಕ್ಯಾನ್ಸರ್ ಆಸ್ಪತ್ರೆಯಲ್ಲಿ ಇವರುಗಳಿಗೆ 'ಭರವಸೆ ಕಳೆದ' ರೋಗಿಗಳೆಂದು ನಿರಾಕರಿಸಲಾಗಿತ್ತೆಂಬುದು ಗಮನಿಸಬೇಕು. ಈ ಪ್ರಕರಣಗಳೆಂದು ಅಧ್ಯಯನ ಯೋಗ್ಯವಾಗಿದ್ದು, ಇದು ಕ್ಯಾನ್ಸರ್ ರೋಗಿಗಳಿಗೆ ಅಪಾರ ಶಕ್ತಿ ನೀಡುತ್ತದೆ, ಡಿಎನ್‌ಆರ್‌ಸಿಯಲ್ಲಿನ ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ಭರವಸೆ ಹೊಂದಲು ಹಾಗೂ ನಿರ್ಧಾರದಿಂದ ಬದುಕಿ ಭಯ ದೂರ ಮಾಡಲು, ಕ್ಯಾನ್ಸರ್‌ನೊಂದಿಗೆ ಸೆಣಸಿ, 'ಕ್ಯಾನ್ಸರ್ ವಿಜೇತ'ರೆಂದೆನಿಸಲು ನೆರವಾಗುತ್ತದೆ.

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

ಈ ರೋಗಿಗಳಲ್ಲಿ ಯಕೃತ್ತು, ಮೆದುಳು, ಶ್ವಾಸಕೋಶ, ಪ್ರಾಸ್ಟೇಟ್ !
ಮೆದೋಜೀರಕ ಗ್ರಂಥಿ. ಮೂತ್ರಪಿಂಡ. ಸೆರ್ವಿಕ್ಸ್‌ಗಳ ಮೆಟಾಸ್ಟಾಸಿಸ್‌ಗೆ
ಒಳಗಾದವರೂ ಇರುವರಲ್ಲದೇ! ಅವರು ಕ್ರಮೇಣ ನಿರಾಶೆಗೆ ಪೂರ್ತಿ
ಬಲಿಯಾಗಲಿದ್ದವರೂ ಇದ್ದರು ದೇಶದ ಒಂದಲ್ಲೊಂದು ಕ್ಯಾನ್ಸರ್ ಆಸ್ಪತ್ರೆಯಲ್ಲಿ
ಇವರುಗಳಿಗೆ ಭರವಸೆ ಕಳೆದ. ರೋಗಿಗಳೆಂದು ನಿರಾಕರಿಸಲಾಗಿತ್ತೆಂಬುದು
ಗಮನಿಸಬೇಕು ಈ ಪ್ರಕರಣಗಳೆಂದು ಅಧ್ಯಯನ ಯೋಗ್ಯವಾಗಿದ್ದು ಇದು
ಕ್ಯಾನ್ಸರ್ ರೋಗಿಗಳಿಗೆ ಅಪಾರ ಶಕ್ತಿ ನೀಡುತ್ತದೆ. ಡಿಎನ್‌ಆರ್‌ಸಿಯಲ್ಲಿನ
ಚಿಕಿತ್ಸೆಯಲ್ಲಿ ಭರವಸೆ ಹೊಂದಲು ಹಾಗೂ ನಿರ್ಧಾರದಿಂದ ಬದುಕಿ ಭಯ
ದೂರ ಮಾಡಲು, ಕ್ಯಾನ್ಸರ್‌ನೊಂದಿಗೆ ಸೆಣಸಿ. ಕ್ಯಾನ್ಸರ್ 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

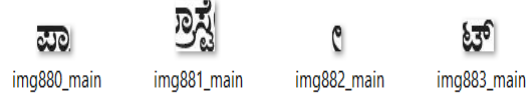
Word

Imperfect Character Segmentation

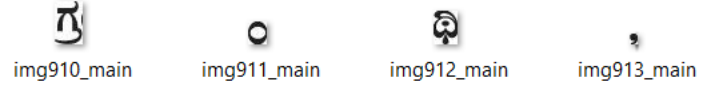
ಶ್ವಾಸಕೋಶ,



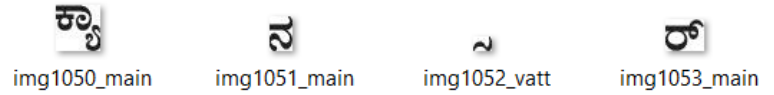
ಪ್ರಾಪ್ತೇಟ್



ಗ್ರಂಥಿ,



ಕ್ಯಾನ್ಸರ್



ಯೋಗ್ಯವಾಗಿದ್ದು,



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
‘ಭರವಸೆ		

“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 ಚೇತನಾಡುತ್ತಾಹದಿ
ನವೀನ ಜೀವನ ಚಿಮ್ಮುತಿದೆ !

ಅಭಿನವ ಮಧುಕೋಕಿಲ ಕಲಾಕಂಠದಿ
ಸ್ವರಸುರ ಚಾಪಗಳುಣ್ಣುತಿವೆ !
ಶ್ಯಾಮಲ ಕಾನನ ಸುಮ ಸಮ್ಮೇಳದಿ
ಇಂಚರ ಸಾಸಿರ ಪೊಣ್ಣುತಿವೆ!

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

ನಾಡಿನ ಪುಣ್ಯದ ಪೂರ್ ದಿಗಂತದಿ
ನವ ಅರುಣೋದಯ ಹೊಮ್ಮತಿದೆ!
ಚೆರನೂತನ ಚೇತನಾಡುತ್ತಾಹದಿ
ನವೀನ ಜೀವನ ಚೆಮ್ಮತಿದೆ

ಅಭಿನವ ಮಧುಕೋಕಿಲ ಕಲಾಕಂಠದಿ
ಸ್ವರಸುರ ಚಾಪಗಳುಣ್ಣತಿವೆ
ಶ್ಯಾಮಲ ಕಾನನ ಸುಮ ಸಮೇನಳದಿ
ಇಂಚರ ಸಾಸಿರ ಪೊಣ್ಣತಿವೆ

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	೦ (Anuswara)	0 (English numeral zero)
ಚೆರನೂತನ	ಚಿ	ಚೆ
ಪೊಣ್ಣತಿವೆ!	೬	೩

In the 1st example, Kannada Anuswara is mispredicted as English numeral “zero”. In the 2nd example, akshara “chi” is mispredicted as “che”. These are understandable since the corresponding

characters have lots of similarity. In the 3rd 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