**International Journal of Information Retrieval Research**
Volume 12 • Issue 1

# An Improved Text Extraction Approach With Auto Encoder for Creating Your Own Audiobook

Shakkthi Rajkumar, Anna University, India
Shruthi Muthukumar, Anna University, India
Aparna S. S., Anna University, India
Angelin Gladston, Anna University, India

## ABSTRACT

As we all know, listening makes learning easier and is more interesting than reading. An audiobook is a software that converts text to speech. Though this sounds good, the audiobooks available in the market are not free and feasible for everyone. Added to this, that these audiobooks are only meant for fictional stories, novels, or comics. A comprehensive review of the available literature shows that very little intensive work was done for image-to-speech conversion. In this paper, the authors employ various strategies for the entire process. As an initial step, deep learning techniques are constructed to denoise the images that are fed to the system. This is followed by text extraction with the help of OCR engines. Additional improvements are made to improve the quality of text extraction and post processing spell check mechanism are incorporated for this purpose. The result analysis demonstrates that with denoising and spell checking, the model has achieved an accuracy of 98.11% when compared to 84.02% without any denoising or spell check mechanism.

## KEYWORDS

Auto Encoder, Denoising, OCR, Spell Check, Text Extraction, Text to Speech

## 1. INTRODUCTION

The present era has witnessed technology etching upon every aspect of our life. Be it, online transactions, reservations and what not, technology has paved its path into our lives. In the current pandemic situation, online education has undergone a massive transition. This is where our audiobooks will be tremendously helpful, especially for school, college students and professors (Noman et. al., 2016). There are a lot of existing systems, including Audible by Amazon, AppleBooks by Apple. Though this sounds good, the existing audiobooks available in the market are not free and feasible for everyone. Not only that, these audiobooks are available only for some prescribed books such as comics or novels. If you take a look at the market, there are a number of services offered by different companies starting from Amazon Audible to iTunes and so on. But in all these services, a narrator reads the story which the user listens to. This involves a lot of human work, and cost.

Hence, we have decided to come up with an audiobook that can be used for any printed material including books, school textbooks as well as storybooks, research papers, newspapers, documents, a

DOI: 10.4018/IJIRR.289570

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

single sheet of text or multiple pages wherein users have to just upload the printed information and enjoy listening to the content. Moreover, this is free of cost, and therefore can be used by anyone. Further, we use an auto-generated voice that translates the text to voice which saves a lot of labour.

The proposed framework has 3 important modules. Preparing clean and clear images for the recognition engines is often taken for granted as a trivial task that requires little attention. However, this step undoubtedly influences the overall performance of the system. The first module involves denoising the images that are fed by the user. For this purpose, Mayank et. al., (2020) have used CNN architecture for separating the foreground and background. This work acknowledges the problems of restoration of the useful content of handwritten documents and reconstruction of the 'most likely' appearance of the original documents. Jose et. al., (2005), Lavanya et. al., (2020) and Chunwei et. al., (2015) have discussed neural networks approach for cleaning the dirty documents.

The second step is text extraction. Text Extraction is useful in information retrieving, searching, editing, documenting, archiving or reporting of image text. Jian et. al., (2013) have tried to find a new way which can comprehensively utilize existing methods to detect and extract text from born-digital image. Cartic et. al., (2012) uses edge detection methods. Similarly, we plot bounding boxes and identify the Region of Interest, ROI.

Following this, we have designed post processing steps to identify the errors and suggest suitable alternatives. However, the basic outline for building this work emerged from the extensive work undertaken by Sasirekha et. al., (2013). A different approach is used to extract text from newspapers, wherein the user has the choice to manually select the columns he wishes to. The underlying analysis will provide a detailed review of each of the modules with the outcome of each step.

## 2. RELATED WORK

Mayank Wadhwani et al. (2020) propose a CNN model that is used to restore the images and separate foreground and background. Here, both background and foregrounds are restored in parallel using two neural networks with similar architecture. Finally, restored foreground and background images are combined to reconstruct (expected) original document image. We propose a similar auto encoder architecture for our system but we do not restore the background since our system is not interested in background details. The authors state that least distinguishable difference between intensities (i.e., contrast) of two adjacent regions depends on the intensity of the neighborhood. However, this perceptual criteria is not reflected in widely used metrics like Euclidean distance employed in common clustering and classification algorithms. So to facilitate text extraction we enhance the grayscale image. In our proposed system too, we enhance the text extraction by performing binarization on the input images so that the OCR engine can extract text clearly.

Jose Luis Hidalgo et al. (2005) has described a generic cleaning and enhancing system for automatic form processing using neural networks. It takes clean and simulated noisy images to train and select the best neural network. Subjective and objective evaluations of the cleaning method show excellent results to clean forms with printed grey-areas to indicate where to fill in the information. The work described here consists of filtering the background noise caused mainly by grey rectangles used as guiding rulers. The proper elimination of these rectangles makes it possible to use this approach in the design of forms to be used by handwritten recognition systems, which is much cheaper than other approaches.

Chunwei Tian et al. (2015) have discussed various deep learning techniques for image denoising. The authors have used dilated convolution methods along with mining edge information as this has achieved significant state-of-art results. Different transfer learning methods such as CBDNet, FFDNet and their corresponding results are compared using metrics such as PSNR. The author emphasizes the need of data augmentation such as performing flipping, adding blur noise to images. Nevertheless, we have performed data augmentation to add upon nearly 1000 images to the existing dataset for our model. Also, since the available dataset did not contain a lot of images, we added our own images

collected from old materials. Since ground truth images were not present for them, we had to clean the images using some latest editing tools so that supervised learning can be performed.

Cartic et. al., (2012) used a baseline system called the LA-PDFText that is designed as a precursor for further improvements to the block detection, classification and text extraction stages. Here, the author has worked in a 3 stage process: to detect text plots using spatial layout processing to locate and identify blocks of contiguous text, classify boxes into categories and connect the extracted box in the correct order. On a similar note, our text extraction can also be applied for PDF files, say research papers. Since research papers contain information such as journal name, year of publishing, URLs, post processing the text extracted yielded better accuracy.

Vibhakti V. Bhaire et al. (2015) put forth an algorithm that is based on trie data structure which is used to store all the possible words in the English language. Each complete English word has an optional integer value related with it. A trie can be seen as a fatalistic fixed automation without iterations. Each finite language is generated by a trie automation, and each trie can be abstracted into a DAFSA. The words are looked up in this dictionary and if they are not present, they are categorized as erroneous words. The author mentions that one of the disadvantages is that proper nouns may be identified as spelling mistakes, proper names of persons or places, as they are not usually found in a traditional dictionary. We have attempted to resolve this by segregating proper nouns, names of people, places, acronyms, URLs so that they are not wrongly classified.

Noman Islam et al. (2016) employ various pre-processing techniques before OCR is applied. An OCR is not an atomic process but comprises various phases such as acquisition, pre-processing, segmentation, feature extraction, classification and post-processing. An important part of pre-processing is to find out the skew in the document. Different techniques for skew estimation include: projection profiles, Hough transform, nearest neighborhood methods. Hough transformation is used to isolate features of particular shape. This is used for identifying edges, columns, rows in tables.

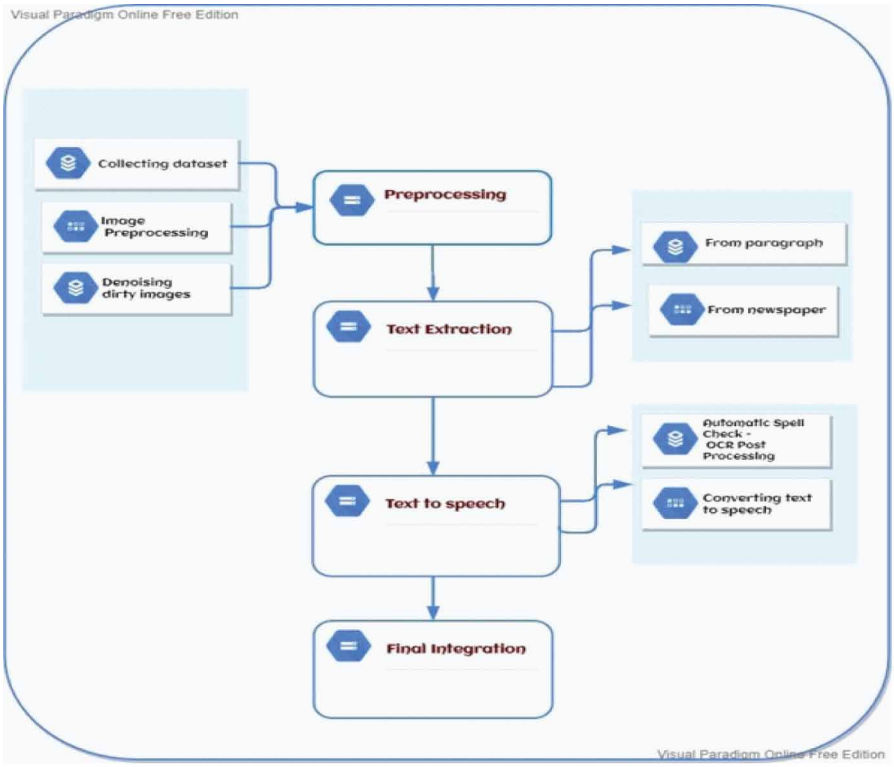
In this paper, our main contributions are: the proposed system adopts various pre-processing techniques such as de-skewing rotated images, finding contours and extracting spatial information. The feature that is new in comparison to this work is post-processing OCR extracted text. We have incorporated mechanisms to identify wrong words and suggest possible words. The work done by Ajinkya et. al., (2013) is similar to this work. The unique part in this work is that we have incorporated post processing work that needs to be done after text extraction from OCR and also extended this facility for newspapers too. Since, content in newspapers are arranged column wise, it is difficult to extract text. Here, we allow users to select their ROI which in turn will be fed to the system to get the audio. Also, we do spell check which is not incorporated in the work mentioned.

## 3. SYSTEM DESIGN

This section discusses in detail the design details of the proposed system creating the audiobook. Figure 1 presents the overall block diagram. The block diagram depicts clearly the main blocks of the proposed audiobook creation, namely preprocessing, text extraction, text to speech and final integration.

Data preprocessing is one of the most important steps which is not given much importance. This includes collecting dataset, image pre-processing and denoising dirty images. The dataset collected contains only 144 images and hence to make our system stable, we add a couple of our own images from old newspapers, story books, and text books. The dataset should contain images taken from different perspectives with varied rotation, lighting. For this purpose, data augmentation is done. This is followed by deskewing anddenoising the images using 3 techniques- auto encoders, median filtering and adaptive thresholding. Text Extraction is the second module which is done from both paragraphs and newspapers. In the case of newspapers, the user has the choice to select the ROI and crop images. Text extraction is done using PyTesseract. The third module includes post processing followed by text to speech conversion. Post processing occurs in 2 steps. Firstly, we need to detect

**Figure 1. The overall block diagram of the system**



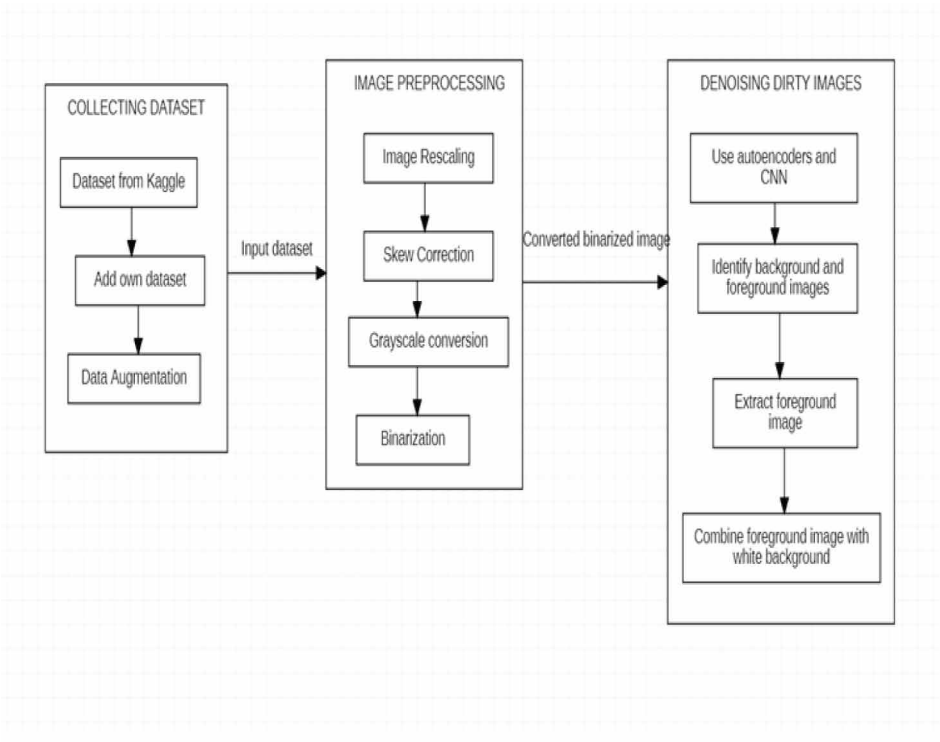

the erroneous words and then correct those using suitable alternatives. After performing spell check, Google Text to Speech API is used to convert the corrected text to audio and store it as an .mp3 file. We finally create an E2E application by integrating all the modules using Flaskframework.

## 3.1 Pre-Processing

**Description:** This preprocessing involves adding our own images to the existing dataset and applying data augmentation techniques like rotation of images by a random angle, adding noise to images and blurring as shown in Figure 2. This is followed by skew correction and binarization which is then fed to one of the denoising methods. Auto encoders, Median Filtering, Adaptive Thresholding are the three techniques that are used to denoise images. In all these techniques, we separate foreground from background and extract only the foreground image with white background.

Auto encoders are an unsupervised learning technique in which we leverage neural networks for the task of representation learning. Specifically, we'll design a neural network architecture such that we impose a bottleneck in the network which forces a compressed knowledge representation of the original input. If the input features were each independent of one another, this compression and subsequent reconstruction would be a very difficult task. The model as shown in Figure 3 is simulated using the following parameter settings to obtain a very efficient system:

**Figure 2. Flow diagram for preprocessing**




- Number of hidden layers: 4
- Activation functions: ReLU
- Number of epochs:100
- Batch size: 4
- Loss function: Mean Square Error
- Optimizer: Adam
- Metrics: Mean Absolute Error

## 3.2 Text Extraction

**Description:** Text Extraction as shown in Figure 4 is carried out from both paragraphs and newspapers. The first step before extracting text is to rotate the denoised image back to 0° because in the deskewing part, the images could have been rotated to any multiple of 90° (0°, 90°, 180°, 270°). After rotating the image back properly, text extraction using PyTesseract happens. Each word is extracted by identifying the region of interest. The top left coordinates along with height and width for each word are determined and a bounded box is drawn. This is the initial level of text extracted from OCR. Since extracting text from newspapers is a tedious process, we manually crop the columns and order it for text extraction. This cropping includes reset (click R), copy (click C) and quit (click Q) options. Once the columns are cropped, the text is extracted by the OCR in the same fashion.

**Figure 3. Auto encoder**

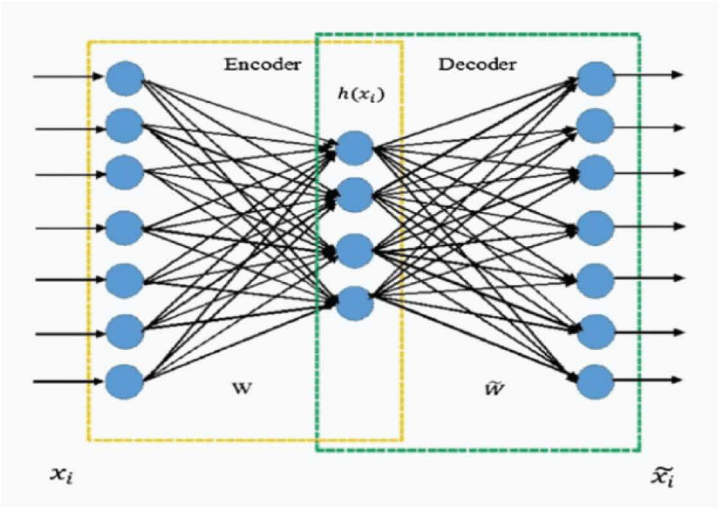
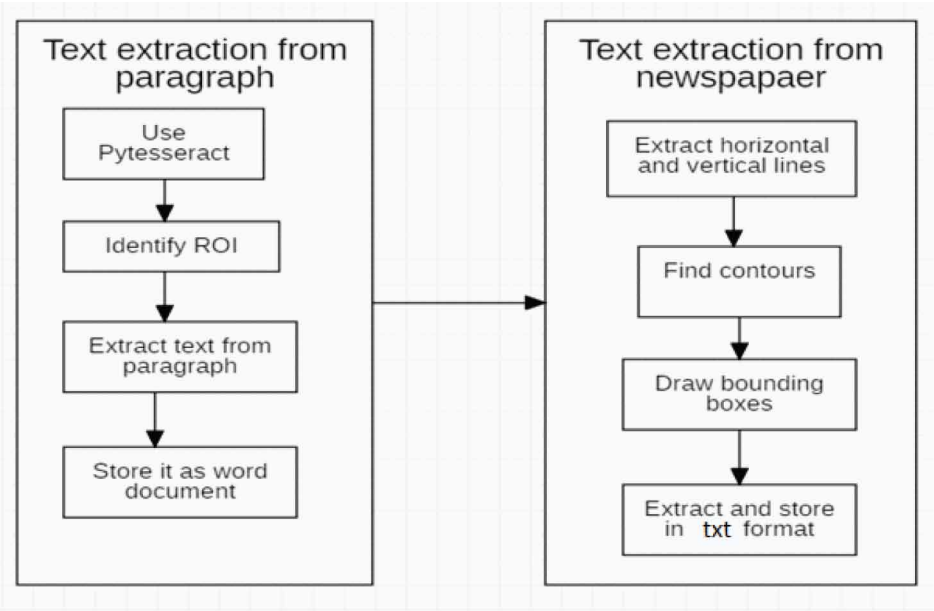



**Figure 4. Flow diagram for text extraction**



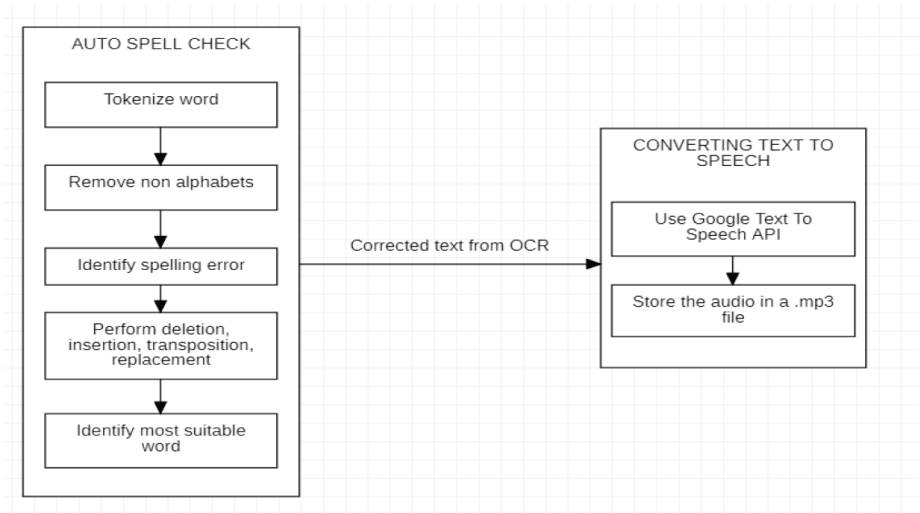
### 3.3 Text to Speech

**Description:** Post processing of OCR is one of the most important feature that normal text extraction from OCRs miss out on. Before applying Spellchecker methods as shown in Figure 5, there are some preprocessing steps to be done:

a. Identifying proper nouns, and capitalized words as they may indicate a person's name or important keywords in case of research paper.
b. Identifying URLs.
c. Identifying e-mail ids as these must not be marked as incorrect words.
d. Remove unnecessary punctuation marks and special symbols.

Words are then looked up in the NLTK dictionary and if a word is not present it is marked as incorrect. To differentiate between incorrect and correct words, we override the incorrect words with the word "[MASK]". All the words are tokenized and passed to the predictor. The predictor takes the [MASK] ed words and replaces them with a suitable word. Suggestions are done based on Levenshtein distance which performs insertion, deletion, transposition, replacement based on the number of characters that must be edited. The corrected text is converted to audio using the Google Text to Speech API and stored as an mp3 file.

**Figure 5. Flow diagram for text to speech**




As part of final integration the auto encoder CNN model is stored as a tensorflow model and this can be used for further testing. All the modules are integrated into one and an E2E application is created with Flask framework.

## 4. EXPERIMENTAL RESULTS

This section presents the experimental details, the results obtained in the experiments. The results obtained for data augmentation, denoising, text extraction from paragraph, text extraction from newspaper and auto spell check.

### 4.1 Data Augmentation

#### *4.1.1 Rotation (Either Clockwise or Anticlockwise Rotation)*

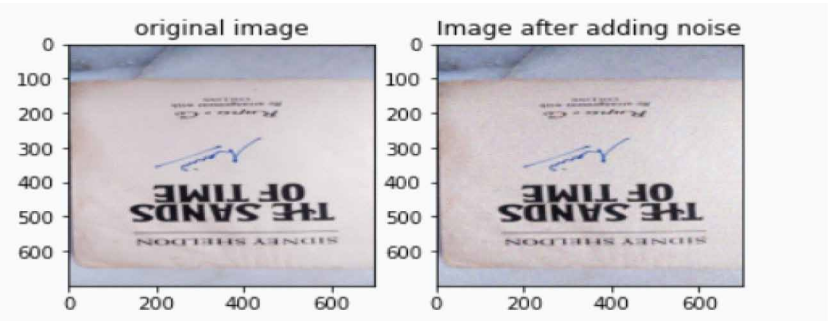
In Figure 6, the image (original image) is rotated in the anticlockwise direction over an angle of 60°.

**Figure 6. Augmentation by anticlockwise rotation- Original image(left), Augmented image (right)**

#### *4.1.2 Adding Noise to the Image*

In Figure 7, the image (original image) is added with noise to generate the augmented image.

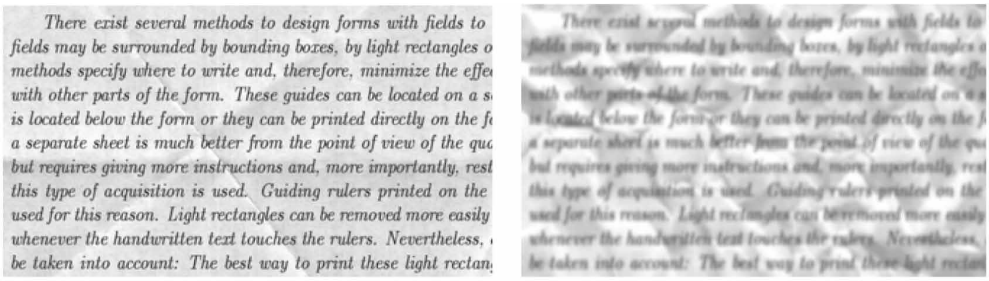
**Figure 7. Augmentation by adding noise- Original image (left), Augmented image (right)**




#### *4.1.3 Blurring the Image*

In Figure 8, the image (original image) is blurred to generate the augmented image.

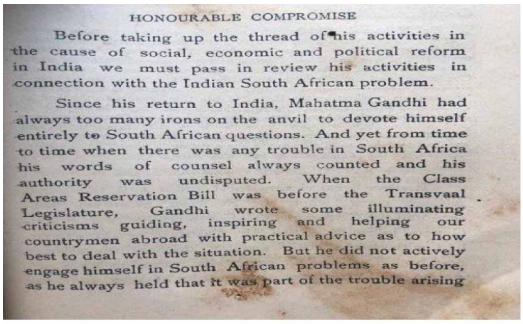
**Figure 8. Augmentation by blurring the image- Original image (left), augmented image(right)**



## 4.2 Denoising

In the figures 9 and 10, the image is denoised using the 3 methods and we find that median filtering is the best method for this image.

**Figure 9. Original image to be denoised**

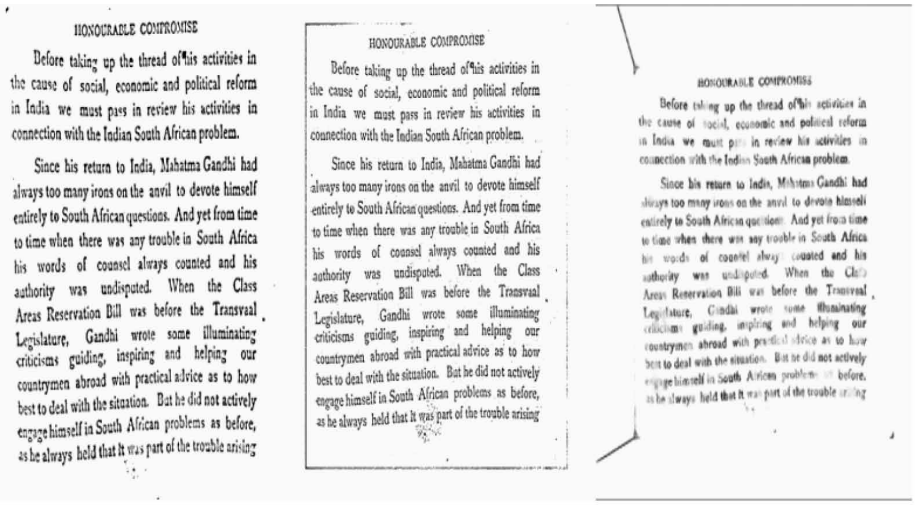


HONOURABLE COMPROMISE

Before taking up the thread of his activities in the cause of social, economic and political reform in India we must pass in review his activities in connection with the Indian South African problem.

Since his return to India, Mahatma Gandhi had always too many irons on the anvil to devote himself entirely to South African questions. And yet from time to time when there was any trouble in South Africa his words of counsel always counted and his authority was undisputed. When the Class Areas Reservation Bill was before the Transvaal Legislature, Gandhi wrote some illuminating criticisms guiding, inspiring and helping our countrymen abroad with practical advice as to how best to deal with the situation. But he did not actively engage himself in South African problems as before, as he always held that it was part of the trouble arising

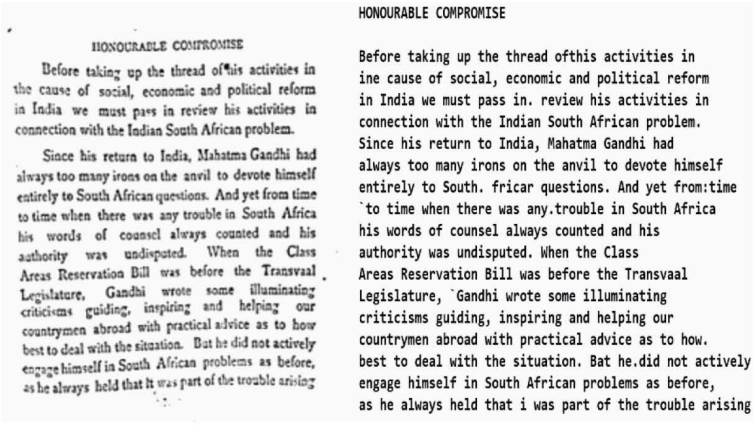
**Figure 10. Denoised image using median filtering (left), adaptive thresholding (middle) and auto encoders (right)**



## 4.3 Text Extraction From Paragraph

In figure 11, the denoised image (using median filtering) is fed to the OCR to get the corresponding extracted text.

**Figure 11. Denoised image (left) and the corresponding extracted text (right)**



HONOURABLE COMPROMISE

Before taking up the thread of his activities in the cause of social, economic and political reform in India we must pass in review his activities in connection with the Indian South African problem.

Since his return to India, Mahatma Gandhi had always too many irons on the anvil to devote himself entirely to South African questions. And yet from time to time when there was any trouble in South Africa his words of counsel always counted and his authority was undisputed. When the Class Areas Reservation Bill was before the Transvaal Legislature, Gandhi wrote some illuminating criticisms guiding, inspiring and helping our countrymen abroad with practical advice as to how best to deal with the situation. But he did not actively engage himself in South African problems as before, as he always held that it was part of the trouble arising

HONOURABLE COMPROMISE

Before taking up the thread ofthis activities in
ine cause of social, economic and political reform
in India we must pass in. review his activities in
connection with the Indian South African problem.
Since his return to India, Mahatma Gandhi had
always too many irons on the anvil to devote himself
entirely to South. fricar questions. And yet from:time
`to time when there was any.trouble in South Africa
his words of counsel always counted and his
authority was undisputed. When the Class
Areas Reservation Bill was before the Transvaal
Legislature, `Gandhi wrote some illuminating
criticisms guiding, inspiring and helping our
countrymen abroad with practical advice as to how.
best to deal with the situation. Bat he.did not actively
engage himself in South African problems as before,
as he always held that i was part of the trouble arising

## 4.4 Text Extraction From Newspaper

In the figure 12, the newspaper clipping is cropped manually column wise and copy it (by clicking C) for extraction using OCR.

**Figure 12. Manual cropping of the newspaper clip**

## 4.5 Auto Spell Check

In the figures 13 and 14, from the extracted text the incorrect words are identified and masked and then replaced with the correct words (by checking with the dictionary) to generate the corrected text document.

**Figure 13. Masked document (left), list of incorrect words (middle) and list of replaced words (right)**

HONOURABLE COMPROMISE
Before taking up the. thread [MASK] activities in
[MASK] cause of social economic and political reform
in India we must pass in. review his activities in
connection with the Indian South African problem.
Since his return. to India Mahatma Gandhi had
always too many irons on the anvil to devote himself
entirely to South. [MASK] questions. And yet [MASK] to time
when there was. [MASK] in South Africa
his words of counsel always counted and his
authority was undisputed. When the Class
Areas Reservation Bill was before the Transvaal
Legislature Gandhi wrote. some illuminating
Criticisms guiding inspiring and helping our
countrymen abroad with practical advice as to how.
best to deal with the situation Bat [MASK] not actively
engage himself in South African problems as before
as he always held that was part of the trouble arising

| S.NO | INCORRECT WORDS |
|---|---|
| 1. | ofthis |
| 2. | inthe |
| 3. | fricar |
| 4. | from:time |
| 5. | any.trouble |
| 6. | he.did |

| S.NO | REPLACED WORD |
|---|---|
| 1. | of this |
| 2. | one |
| 3. | African |
| 4. | foretime |
| 5. | trouble maker |
| 6. | Heidi |

**Figure 14. Corrected text document**

HONOURABLE COMPROMISE
Before taking up the. thread of this activities in
one cause of social economic and political reform
in India we must pass in. review his activities in
connection with the Indian South African problem.
Since his return. to India Mahatma Gandhi had
always too many irons on the anvil to devote himself
entirely to South. African questions. And yet foretime to time
when there was. troublemaker in South Africa
his words of counsel always counted and his
authority was undisputed. When the Class
Areas Reservation Bill was before the Transvaal
Legislature Gandhi wrote. some illuminating
Criticisms guiding inspiring and helping our
countrymen abroad with practical advice as to how.
best to deal with the situation Bat Heidi not
actively engage himself in South African problems as before
as he always held that was part of the trouble arising

**Figure 15. Final audio file**

Hence, the audio for the corrected text document is generated using Google API and is played in the front end.

## 5. RESULT ANALYSIS

This section presents the result analysis of the audiobook creation using the proposed approach. The various evaluation metrics used in the experimental evaluation are Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE). The PSNR and MSE values obtained are tabulated in Table 1 along with the original image as well as the denoised image.

**Table 1. Evaluation metric for quality of images**

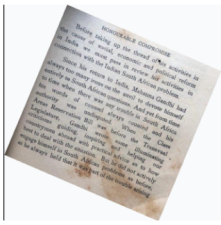
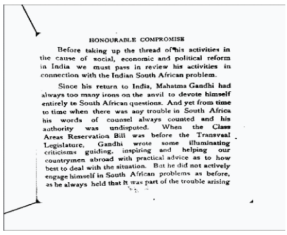
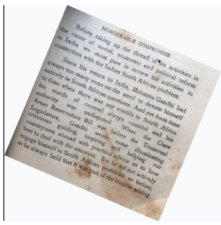
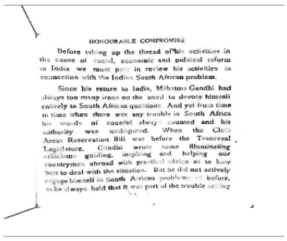
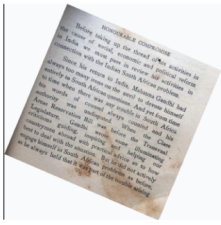
| Original Image | Denoised Image | Parameters | |
|---|---|---|---|
| | | Peak Signal to Noise Ratio (PSNR) | Mean Square Error (MSE) |
|  |  Median filtered image | 31.811116 dB | 42.851765 |
|  |  Auto encoded image | 14.653431 dB | 2227.087213 |
|  | Adaptive thresholded image | 31.809817 dB | 42.864584 |

Image quality assessment is subjective and varies from person to person. Hence, for the same set of images, different denoising algorithms will give different results. The two evaluation metrics we have used in our system to assess the quality are PSNR and MSE.

PSNR is the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of representation and this is measured in logarithmic scale because many signals have wide dynamic range difference. The higher the PSNR, the better the quality of the compressed, or reconstructed image. As we see, for the above 3 images, PSNR for median filtered image is high compared to adaptive thresholding and PSNR for auto encoder is the least amongst the three.

MSE, is nothing but the well-known metric that is calculated as the sum of square of difference between predicted and target image. Here, each of the pixelated values are compared against each other and a low value for MSE must be preferred. Here, we see that the MSE of the median filtered image is the least, which is around 42 while the MSE of the auto encoder image is 50 times that of the other two.

For a better quality image, the model with high PSNR value and low MSE value must be chosen and that varies from image to image. From the above 2 inferences, we can conclude that for the given picture, median filtering is the best choice for denoising as PSNR value is high and MSE is low:

$$\text{Peak Signal to Noise Ratio (PSNR)} = 20 \log_{10} (MAX_f / \sqrt{MSE}) \quad (1)$$

$MAX_f$ = Maximum signal value in the original image

$$\text{Mean Squared Error (MSE)} = 1/n \, \Sigma \, (x_i - h(x_i))^2 \quad (2)$$

$x_i$ = True value
$h(x_i)$ = Predicted value
n = Total number of data points

$$\text{Mean Absolute Error (MAE)} = \Sigma \, |y_i - x_i| / n \quad (3)$$

$y_i$ = Prediction
$x_i$ = True value
n = Total number of data points

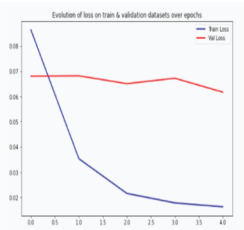
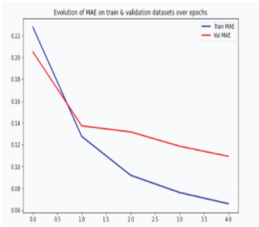
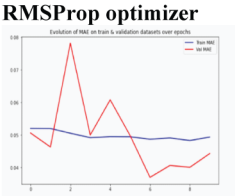
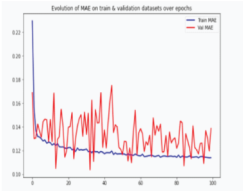
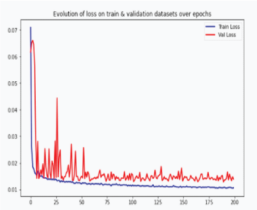
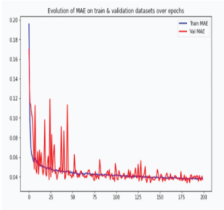
Table 2 tabulates the values of training and validation losses as well as the mean absolute error values for the various models which helps in choosing the best model.

The model is initially trained for 5 epochs with 2 different optimizers, namely Adam and RMSProp optimiser. From the graph, we can infer that though there is significant improvement in loss function, the other metrics (MAE - Mean Absolute Error) have great fluctuations.

Consequently, training the model for a large number of epochs will lead to better results.Adam Optimiser was trained for 100 epochs with a batch size of 4 and patience level of 50. To prevent over fitting, we set dropout to 0.3 and perform batch normalization. From the graph, we can see that the difference between training and validation loss is 0.01 and there are huge fluctuations for MAE throughout 100 epochs. The validation MAE is not at all stable for the entire 100 epochs and it can be inferred that training MAE has converged but so is not the case with validation MAE.

We observed that RMSProp optimiser performed well and hence further increased the no of epochs to 200 in order to train the model efficiently. At the end of 200 epochs, training loss is only about 0.01 and generalization error between training and validation loss is a meagre 0.005. In contrast to the previous model, the RMSProp optimiser has yielded better results with regards to MAE also.

**Table 2. Plots for training and validation loss and mean absolute error for different models**

| Inference | Parameters | |
|---|---|---|
| | **Loss Function** | **Mean Absolute Error (MAE)** |
| Epochs = 5<br>We first train to see how the model evolves for 5 epochs. From the graphs, we can infer that significant accuracy is not achieved and hence we increase the no of epochs |  <br>**RMSProp optimizer**<br> <br>**Adam optimizer** |  <br>**RMSProp optimizer**<br>  <br>**Adam optimizer** |
| Epochs = 100<br>With Adam optimizer, though the generalization loss is not much, we can see fluctuations with MAE which may cause over fitting | <br>**Adam optimizer** |  <br>**Adam optimizer** |
| Epochs = 200<br>We have finally decided to go with the RMS Prop model since here both training and validation loss as well as MAE coincides and the model does not over |  <br>**RMSProp optimizer** |  <br>**RMSProp optimizer** |

Despite having fluctuations in the initial 50 epochs, we see the validation curve of MAE has converged smoothly and coincides with the training curve of MAE. It can hence be concluded that RMSProp optimiser is better in this case than Adam optimiser.

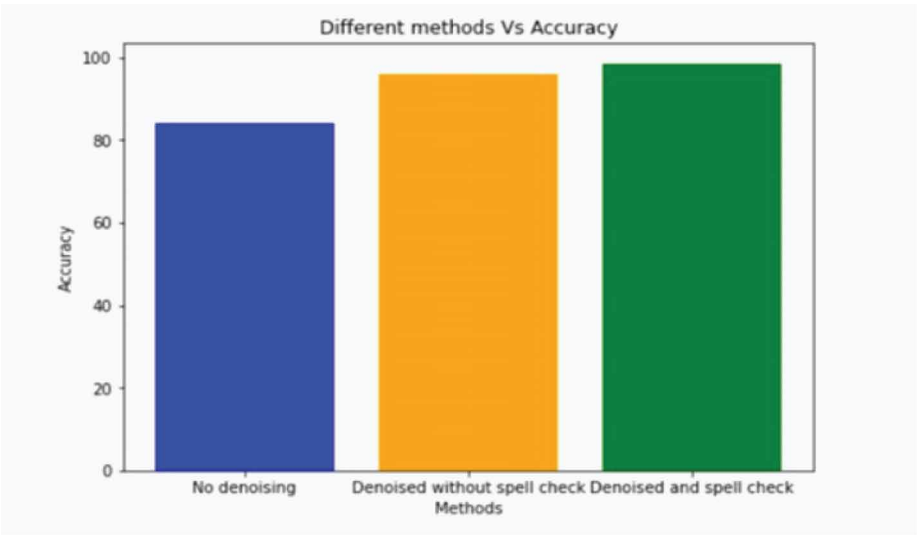
Further, the proposed system has been compared with the existing system as shown in the Table 3. The proposed system introduced the spell check and the Table 3 tabulates the case without spell check as well as the system with spell check. The values indicate 3% increase with spell check which indicates that the proposed audiobook creation performs well with improved accuracy.

Accuracy here is defined as the ratio of no of incorrect words to the total no of words present in the text. From the tabulations, Table 3 as well as Figure 15, which plots the various accuracy values, we can interpret that post processing OCR has helped to achieve even more significant results.

**Table 3. Accuracy of the text before and after spell check correction**

| | No. of Incorrect Words (Total words in given text = 144) | Accuracy |
|---|---|---|
| NO DENOISING | 23 | 84.027 |
| DENOISED WITHOUT SPELL CHECK | 6 | 95.833 |
| DENOISED WITH SPELL CHECK | 2 | 98.611 |

**Figure 16. Accuracy plot to compare with existing and proposed system**




With no pre or post processing methods, the existing text extraction has yielded an accuracy of about 85%. From the bar plot, we can infer that the post processing method i.e. denoised with spell speck gives significant accuracy of about 98.6% compared to 95% with only denoising and no spellcheck.

## 6. CONCLUSION

At times like COVID, when everything is digital, we find it strenuous to stare at the screens all the time. This is where the audiobooks come to the rescue. In contrast to the existing OCR system, our system has achieved significant results. The three different denoising methods helped in better text extraction from OCR. Median filtering method is used in case of research papers, adaptive thresholding methods are used in case to preserve certain images and auto encoders are helpful in removing stains as background. The proposed system will not only build more accurate text extraction but also does post processing spell checking which is generally not present. Conclusively, the proposed system

is by itself an innovation to the market place and the pre-processing and post-processing methods we have applied has made the system even more efficient and effective. In future, we can provide a better mechanism for newspapers such that the columns are automatically detected instead of manual selection. For now, we have restricted the language to only English. We can further extend the feature to other languages such as Tamil.

## REFERENCES

Bhaire, Jadhav, Pashte, & P.G. (2015). Spell Checker. *International Journal of Scientific and Research Publications, 5*(4).

Domale, Padalkar, & Parekh. (2013). Printed Book to Audio Convertor for Visually Impaired. In *2013 Texas Instruments India Educators' Conference*. IEEE. doi:10.1109/TIIEC.2013.27

Etoori, P., Chinnakotla, M., & Mamidi, R. (2018). Automatic Spelling Correction for Resource-Scarce Languages using Deep Learning. *Proceeding of ACL 2018*, *Student Research Workshop*. doi:10.1109/TIIEC.2013.27

Hidalgo, Espana, Castro, & Perez. (2005). *Enhancement and Cleaning of Handwritten Data by using Neural Networks*. Springer-Verlag Berlin Heidelberg.

Islam, Islam, & Noor. (2016). A Survey on Optical Character Recognition System. *Journal of Information & Communication Technology, 10*(2). doi:10.1109/TIIEC.2013.27

Natei, Viradiya, & Sasikumar. (n.d.). Extracting Text from Image Document and Displaying its Related Information. *K.N. Natei Journal of Engineering Research and Application.*

Ramakrishnan, Patnia, Hovy, & Burns. (2012). Layout-aware text extraction from full-text PDF of scientific articles. *Source Code for Biology and Medicine 2012*.

Sasirekha & Chandra. (2013). Text Extraction from PDF Document. *Amrita International Conference of Women in Computing (AICWIC'13)*.

Tian, Fei, Zheng, Xu, Zuo, & Lin. (2015). Deep Learning on Image Denoising. *Robust text extraction in images for personal event planner, Eleventh ICCNT 2020*.

Wadhwani, M., Kundu, D., Chakraborty, D., & Chanda, B. (2020). Text Extraction and Restoration of Old Handwritten Documents. arXiv:2001.08742v1 [cs.CV]

Zhang, Cheng, Wang, & Zhao. (2013). Research on the text detection and extraction from complex images. *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*.

*Shakkthi Rajikumar is an under graduate student of the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include image processing, and information text processing.*

*Shruthi Muthukumar is an under graduate student of the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include image processing, and text processing.*

*Aparna S. S. is an undergraduate student of the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include image processing and information retrieval.*

*Angelin Gladston is working as Associate Professor at the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include software engineering, software testing, image processing, social network analysis and data mining.*