Development of VisionAid: Real-Time Document and Natural Scene Text Recognition with Text-to-Speech Assistance for Visually Impaired Individuals using YOLO v9

De Guzman, Chen Abriel D.

Caparas, Allen Jay

Questeria, Alven C.

Chapter 1

THE PROBLEM AND ITS BACKGROUND

Introduction

Inspired by the profound difficulties faced by visually impaired individuals in accessing textual information, the group of students known as Oculi (Latin for “eye”) began developing the app VisionAID Real-Time Image-to-Speech Assistance for Visually Impaired Individuals using Computer Vision. Driven by a shared ambition to empower those with visual impairments through innovative, accessible technology, Oculi’s VisionAid seeks to address the accessibility challenges in the Philippines. Many people in the Philippines have bad eyesight, according to The 2019 Philippine National Blindness Survey by the Department of Health (DOH) reveals that about 1.98% of the population in the Philippines.Nonetheless, the increase in the use of mobile phones, which is now considered a modern basic need for the lowest household income bracket, has redefined accessibility. VisionAid is working to address this very challenge by utilizing the availability of mobile devices to offer a practical and cost-effective solution for the visually impaired individuals.

According to Lancet Global Health Commision highlights that these prescription glasses lack affordability and are not often considered a luxury for those who live in low income countries like the Philippines(Burton, 2020). Existing solutions are costly, intricate, or unsuitable for communities of users as a result of their socio-economic backgrounds. Additionally, there is a current absence of mobile applications that utilize real-time, mobile-based image recognition, text extraction and speech synthesis that is also effective and user-friendly. This emphasizes the fact that there is a need to develop a solution that can be implemented on ordinary and commonly used mobile devices with the capability to offer accurate and reliable support in real situations.

VisionAid’s goal is to provide a real time text-to-speech that can assist not only the visually impaired individual but also it creates an innovative and accessible

environment. With these improvements in accessibility can increase independence and confidence among the users. VisionAid seeks to help the visually impaired individuals read text documents and natural scenes to improve their quality of life (QoL).

Background of the Study

Vision is one of the important senses we human have as it is the primary way to receive information in the environment. It allows us to see colors, shapes, distance and movements as it is crucial for navigating through physical spaces. Visual input is one of the ways for learning, especially for children to help develop their motor skills and it is effective for communication that can recognize facial expressions, body language and other non-verbal signals.

In order to be classified as normal vision, you should have a 20/20 vision, meaning you should be able to see an object clearly from 20 feet away (Cleveland Clinic, 2022). There are multiple types of visual impairment such as cataracts, uncorrected fractional error (nearsighted, and farsighted), glaucoma, and maculopathy. According to the 2019 Philippine National Blindness Survey (DOH), 1.98% of the population in the Philippines has vision impairment or blindness. This equates to 1.1 million Filipinos having cataracts, 400,000 with an uncorrected fractional error, 300,000 with glaucoma, and 200,000 with maculopathy (Villanueva, 2022).In order to be classified as PWD for vision in the philippines you need to have the following, 20/70 vision, field of vision is less than 20 degrees wide in the better eye, vision cannot be improved by eye glasses, medication or surgery (ROQUE Eye Clinic, 2024).

According to the Philippine Statistics Authority, the poverty incidence among the population in the first semester of 2023, or the proportion of poor Filipinos whose per capita income is insufficient to cover their basic food and non-food needs, was estimated at 22.4 percent, or 25.24 million Filipinos. Subsistence incidence among Filipinos, or the proportion of Filipinos whose income is insufficient to cover even basic

food needs, was recorded at 8.7 percent, or approximately 9.79 million Filipinos(Philippine Statistics Authority, 2023).

With mobile phones being classified as one of our modern necessities, it’s evident that they play a crucial role in connecting people nowadays. According to the data on mobile user operating systems in the Philippines as of January 2024. 85.19% of all users in the Philippines use Android, whereas 14.31% use iOS (Statista, 2024). This means that Android is used by more people in the Philippines than iPhones.

Given the significant impact of vision impairment on individuals' lives and the financial challenges faced by many in the country, utilizing artificial intelligence (AI) presents a possible solution to this problem. By having a tool designated to cater visually impaired individuals, allowing them to see things that normally cannot. We intend to find a solution in this existing gap, and improve their quality of life.

Objective of the Study

The objective of this study is to develop a real-time document and scene text recognition system with text-to-speech assistance, specifically designed for visually impaired individuals. This study needs to understand the challenges of the visually impaired individuals and provide them with auditory feedback from text documents and natural scenes. This application will have a voice command feature that can help these individuals navigate through the app. It also supports languages like Tagalog, English and Baybayin that are suitable for the people in the Philippines.

Specific Statements of the Study

The research project aimed to achieve the following objectives:

1. Understand the Challenges: Identify the challenges faced by individuals with visual impairments, particularly in accessing written documents and finding effective ways to improve the quality of life.
2. Analyze the Techniques: Evaluate advanced techniques in optical character recognition (OCR) and machine learning to enhance the accuracy and performance of VisionAid.
3. Model Testing: Test and evaluate the VisionAid model using key performance metrics such as Accuracy, F1 Score, Precision, and ISO 25010 to assess software quality and effectiveness.
4. Develop the Mobile Application: Create a mobile application that integrates the machine learning model to enable users to read documents and natural scenes using the device's camera.
5. Application Functional and Non-Functional Requirements: Define and assess both functional (core features such as document reading, scene recognition etc…) and non-functional (performance, security, usability etc…) requirements of the application.
6. User Satisfaction Evaluation: Evaluate user satisfaction with VisionAid using a Likert scale and technology acceptance model (TAM) to assess usability, satisfaction, and effectiveness. and utilizinng ISO 25010

Scope and Limitations of the Study

This study focuses on the visually impaired individuals in the University of Makati. This VisionAid will support languages in the Philippines including Tagalog, English and Baybayin, creating a practical and user-friendly application that enhances independence and accessibility. VisionAid is limited to Tagalog, English and Baybayin, it may also not recognize voices that have a different accent than the accent model is trained on. This application will require the internet for real-time processing of text recognition and for its environmental limitations such as lighting condition, text quality and complexity.

Operational Definition of Terms

Baybayin – An ancient Filipino script used before the Spanish colonization. In the paper, Baybayin is supported as a language, allowing visually impaired individuals to access the ancient script.

Cataracts – A medical condition where there is a cloudy area in the lens, leading to blurred vision. VisionAid aims to assist these individuals by providing auditory output from text and natural scenes.

Convolutional Neural Network (CNN) – A type of deep learning algorithm used to process data and enhance accuracy in image recognition and text extraction.

Connectionist Temporal Classification (CTC) – An algorithm used to train deep neural networks for tasks like speech recognition and handwriting recognition.

Fully Convolutional Network (FCN) – A type of neural network used for tasks such as semantic segmentation, where the goal is to classify each pixel in an image. VisionAid uses FCNs to enhance document layout analysis and text line detection, improving the overall text recognition process.

Glaucoma – A group of eye diseases that can cause vision loss and blindness by damaging the optic nerve. VisionAid aims to assist these individuals by providing auditory output from text and natural scenes.

Maculopathy – A disease related to the central part of the retina, leading to vision loss in the central part of the eye. VisionAid aims to assist these individuals by providing auditory output from text and natural scenes.

Naïve Bayes – A classification algorithm used in related studies to classify data and compare it with other datasets.

Optical Character Recognition (OCR) – A technology that transforms pictures into text. VisionAid employs OCR to recognize and extract text from images captured in real-time.

Region-based Convolutional Neural Network (R-CNN) – A type of deep learning algorithm used for object detection in computer vision.

Scene Image-Text Matching (SITM) – The detection and recognition of scene text from a camera, used for interpreting the context in which text appears within an image.

Support Vector Machine (SVM) – A supervised learning model used for classification and regression analysis. In the paper, it is used to classify and recognize text characters.

Text-to-Speech (TTS) – A technology that transforms text into spoken words. VisionAid uses TTS to provide auditory feedback to visually impaired individuals, enabling them to hear text from both documents and natural scenes.

You Only Look Once (YOLO) – A real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. It is applied in the study for better results and accurate understanding of the model from document reading.

Chapter 2

REVIEW OF RELATED LITERATURE AND STUDIES

It is known for a fact that these types of character recognition had already existed and had already been improved. With this segment, the review of the following literature that had been aforementioned will be done on this chapter and more thorough understanding of the studies will be represented from each of the studies. Finally, the application of these models and algorithms will be the basis of the model that the researchers are planning to build and implement.

Image Character Recognition using Convolutional Neural Networks.

Convolutional Neural Network (CNN) is applied within this study for the better text recognition of the following study. This enhances and does preprocessing of the text from images that had been taken, with the preprocessing applied it had the text recognition increased up to 97.59% with a minimal loss of 6.6% thus concluding the application of the CNN can be applied for the further improvement of the text to speech however further improvements are needed for text that are out of the scenery, thus having trouble for the model to recognize signs that are from a far are the text the authors had problem dealing with as the said concern requires a lot of cleaning.

Classification of Documents Extracted from Images with Optical Character Recognition Methods.

Application of OCR (Optical Character Recognition) method is the utilization of text recognition with machine learning methods so that the text recognition will be able to adapt to different types of handwritten and printed documents as introduced by Aydın, Ö. (2021). in this research, The researchers had used the OCR and applied the Naive Bayes Algorithm to class the the data that the system had been given however when compared to the MODI (Microsoft Office Document Imaging Library) was used, from the given example the OCR system detected 346 characters and 61 words from a

sample text, while MODI detected 351 characters and 62 words. The word match rates were approximately 85.24% for MODI and 88.52% for the OCR method, with both methods achieving a 97.7% character match rate. Finally suggested in order to have higher accuracy for the application of the OCR, it is recommended by the authors to have much more cleaning data to be applied for the images in order to have a much more clear text recognition on the hand written documents, suggesting to use neural networks for the said statement.

An end-to-end Optical Character Recognition approach for ultra-low-resolution printed text images.

This study from Gilbey, J. D., & Schönlieb, C.-B. (2021), approaches OCR (Optical Character Recognition) on ultra-low-resolution images (60 and 75 dpi). Traditional OCR models struggle with such low resolutions, as demonstrated by the difficulty in reading enlarged low-resolution text images compared to higher resolution ones. This challenge is tackled by a new technique inspired by human vision. Multiple methods had been used in order for the cleaning to be applied, such as Nearest-neighbor with application of interpolation followed by Gaussian filtering with different standard deviation, Blurring the images was used to upscale with the application of Gaussian filters to reduce high-frequency noise making edges less distinct and text easier to recognize.

Accuracy metrics that was accepted were the following Character Level Accuracy, Word Level Accuracy which for 60 dpi images: Character error rate reduced by 64% and word error rate reduced by 73% rate and the 75 dpi images had Character error rate reduced by 35% and word error rate reduced by 51%

The proposed methods include upscaling with various interpolation techniques, modifying the Tesseract’s pipeline, and using an ensemble approach to significantly improve the OCR performance on low resolution images. While further research can be

suggested to focus on the fine-tuning for specific text genres and optimizing ensemble methods for even better research.

CNN-RNN BASED HANDWRITTEN TEXT RECOGNITION.

According to Hemanth, Jayasree, Venii, Akshaya, and Saranya (2021), the Handwritten Text Recognition (HTR) system delves into text recognition methods aimed at digitizing both handwritten and machine-generated fonts to address challenges such as varying styles, fonts, and character distortions. The datasets used for training and validation are from IAM Dataset consisting of 100,000 images of handwritten text. In their study, images were changed to grayscale, with boundary boxes drawn for line segmentation and word segmentation techniques applied to break down text lines into individual words. The RNN then extracted relevant sequential information. The application of CTC (Connectionist Temporal Classification) significantly improved accuracy, with the proposed model achieving a word recognition accuracy of 98%, which is notably higher than current benchmarks. The researchers suggest that future work should incorporate hybrid datasets to further address challenges such as broken text recognition and many more.

Attention-Based CNN-RNN Arabic Text Recognition from Natural Scene Images.

According to Butt, Muhammad R. R., Muhammad J. R., Muhammad J. A., and Haris (2021), the application of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for Arabic text recognition from natural scene images utilizes deep learning algorithms to read Arabic text, which is often considered complex due to its unique writing style and format. The researchers employed these neural networks to develop a system capable of reading text while eliminating background scenery, color, and other distractions. Results indicate successful recognition at the character, word, and line levels, surpassing existing methods. Concluded by the study that this also outperforms other models like Deep Belief Networks and Multidimensional Long Short-Term Memory Networks, achieving higher character recognition rate. Sequence

learning techniques enable direct transcription of images, with contextual modeling possible in both forward and backward directions. From the Arabic OCR engine fine reader used for Arabic Character recognition this OCR obtains a 82.4% to 83.26% character recognition rate however the proposed model by the researchers had a model which obtains 98.73%. Given the availability of effective feature learning algorithms, the research community should shift focus towards tackling more challenging tasks such as natural scene recognition.

Optical character recognition system for Baybayin scripts using support vector machine.

From the study of Pino, Mendoza, and Sambayan (2021), a character recognition system was developed for Baybayin scripts using SVM (Support Vector Machine) to recognize both Baybayin and Latin scripts. The preprocessing step converts the original image into binary data. If the image contains only one component, it proceeds with the algorithm; if it includes an accent, the main component and the accent are separated and processed individually. The SVM is then utilized to classify the models, and character recognition is used to classify the characters based on the extracted features. The authors conducted tests on 1,100 randomly chosen images, from the datasets collected from 9,000+ images for Baybayin characters which were taken from the dataset provided by Nogra (2019) in Kaggle. Achieving the following performance metrics: Accuracy of 98.41%, Precision of 98.68%, Recall of 98.45%, and an F1 Score of 98.57%. The proposed OCR system leveraging SVM models demonstrated high performance in both binary and multiclass classification tasks for Baybayin and Latin scripts, indicating a strong capability to accurately classify characters from these scripts. This system holds potential for future research and improvements in script recognition technologies.

Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection.

In their 2023 study, Fateh, Fateh, and Abolghasemi investigate improvements to Optical Character Recognition (OCR) by application of deep learning models like YOLOv3, SSD, Faster R-CNN, and Layout Parser. These models help to classify and extract textual and non-textual regions within documents, providing coordinates and classifications that enhance OCR readability on the documents. A voting system is applied to determine whether a region is textual or non-textual based on the predictions of the said models.

The authors also implemented Text Line Detection (TLD), which optimizes OCR by performing tasks such as angle correction, spacing adjustment, size normalization, and curvature analysis. Preprocessing steps are applied to distinguish between text and non-text content, followed by the implementation of Tesseract-OCR to convert the processed content into more readable formats.

Results Summary**:** The study showed that applying Text Line Detection (TLD) significantly reduced OCR error rates across various datasets. Specifically, the total error rate was reduced from:

- **6.24% to 3.43%** in the ION Dataset,
- **13.88% to 1.52%** in the Arabic Dataset,
- **6.22% to 3.88%** in the Synthetic Dataset.

These improvements demonstrate the effectiveness of TLD in enhancing OCR accuracy, particularly for complex scripts.

DocBed: A Multi-Stage OCR Solution for Documents with Complex Layouts.

From the study of Zhu, Sokhandan, Yang, Martin, and Sathyanarayana (2022), the proposed study on a multi-stage OCR solution for documents with complex layouts aims to enhance OCR performance in complex documents by utilizing a specific sequence of processing steps. The FCN (Fully Convolutional Network) model is employed to perform pixel classification, identifying various segments within the documents, including feature extraction, down-sampling, upscaling, refinement, and classification.

Thus concludes the following:

- SETR-Heuristic provides the best results for text segments with high accuracy and precision.
- SepDet-Geometric delivers the best sequential ordering of text.
- Mask R-CNN and SETR models are effective for detailed layout segmentation, with the Mask R-CNN R50-FPN model being particularly suited for high-accuracy segment classification.
- For applications requiring real-time processing, Mask R-CNN with R50-DC5 backbone is preferred due to its lower inference time.

This study proves that different models excel depending on the specific application requirements, such as accuracy, inference time, and sensitivity to layout classification, making them usable for complex text documents like newspapers.

Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks.

The utilization of the following Convolutional Neural Network (CNN) for the application into OCR to accurately identify and translate handwritten text and printed text documents from various regional languages into a universal spoken language which is English. On the topic of *Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks* The researchers had prepared preprocessing on the images to cutting importance and relevance to the image, gray scale conversion and inverting images, included segmentation is removing borders and splitting text into characters, datasets that had been used for the data preprocessing and training are the MNIST dataset from the Keras library which contains handwritten digits images . With the model that had been applied the recognition accuracy had around 99% accuracy on the MNIST dataset, showing excellent accuracy on handwritten digits. With this result this opens up for a practical application in future programs that will have number recognition.

Baybayin Script word Recognition and Transliteration Using a Convolutional Neural Network

According to Vilvara, R. A., Hammond, D. S. C., Santos, F. M., & Alar, H. S. (2022), The famous writing script of the Philippines was being revived at the time of the making of the following study for cultural preservation and education purposes, as a growing need of tools for potential improvements of learning of the ancient script, the researchers of this study had suggested the use of the Convolutional Neural Network (CNN) based transliteration model capable of converting Baybayin symbols into corresponding Filipino words. The datasets that were used came from 3 sources as indicated within the study, all of these datasets were tailored for character classification, word detection and transliteration. The baybayin character classification that was proposed model of **CNN-Based Transliteration Model** achieved accuracies of 97.4%

and 97.26% from VGG16 based classification, baybayin word detection achieved 96.15%, and the transliteration demonstrated an accuracy of 91.54% on the suggested model. Finally the researchers concluded that the proposed model did achieve promising results, however allow room for improvement as other existing models such as from such as Pino et al.'s system using Support Vector Machines (SVM) with 97.6% accuracy, the use of Levenshtein distance still proves effective for single-word recognition., finally the researchers of this study hopes to contribute further development of Baybayin recognition and transliteration models in the field of computer vision and cultural preservation.

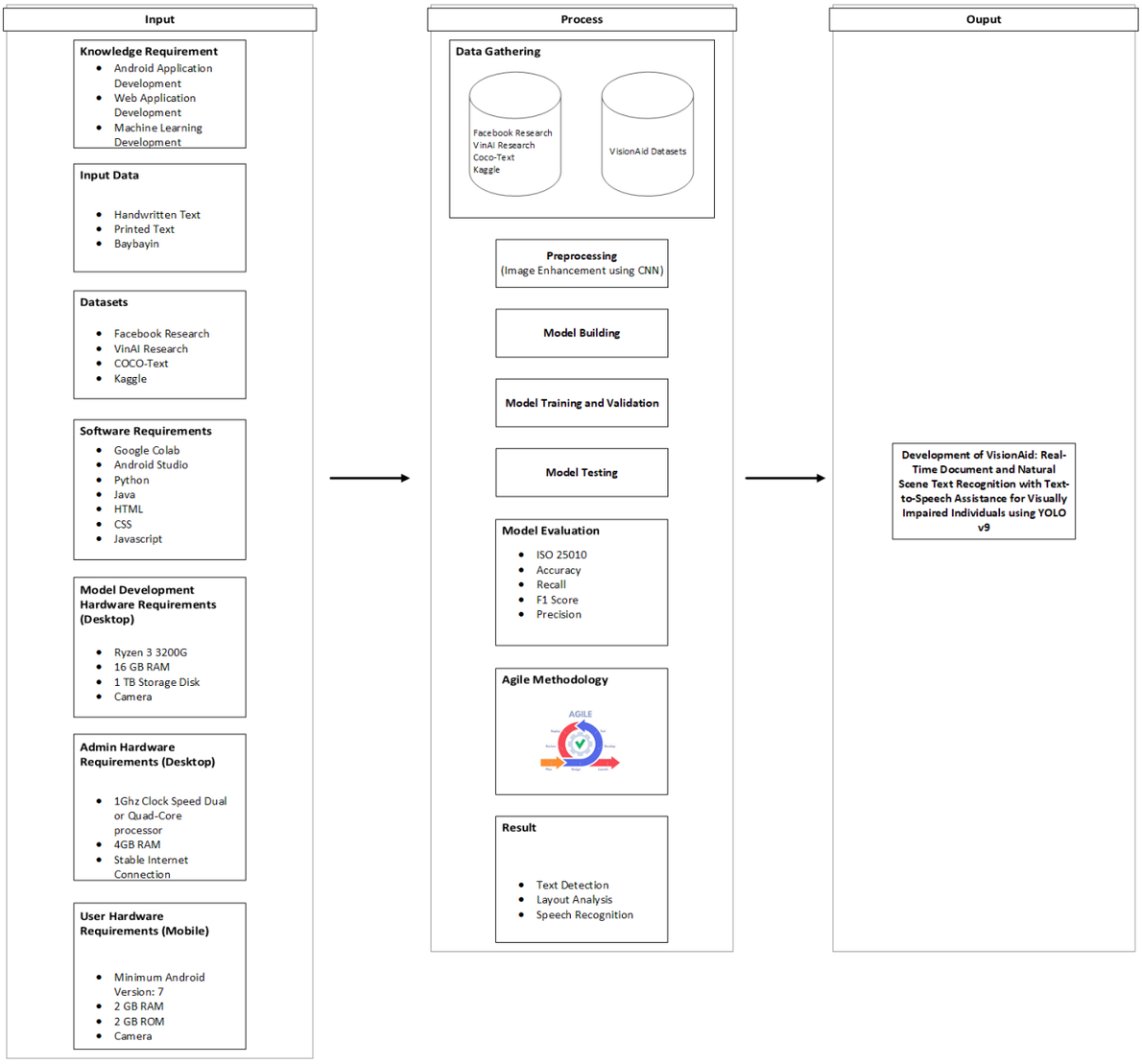
Conceptual Model of the Study

With the following studies mentioned above, the program implies to gather information and collect data from existing structures and have following neural networks applied to the OCR to be developed which is the developmental of the VisionAid that helps secure the accuracy recognition of the text with the application of text recognition, handwritten documents, printed documents, segmentation of the text, ordering, scenery challenges, distance and even the sequence of the complicated text that has different types of writing methods such as arabic and the and the application of the following program towards baybayin writings and the known english language, and finally to implement to Filipino text reader to develop the following document.

The research acknowledges and will be utilizing following data flow with agile application and with following machine learning algorithms aforementioned on the following studies.

The application of convolutional neural network (CNN) is prevalent from all of the following studies with the application of the support vector machine (SVM) for

recognition of baybayin writings, YOLOv9 with application of the following improvement suggested by other studies as enhancing it from different sceneries and segmentation are applied for better result and accurate understanding of the model from the document it is reading.



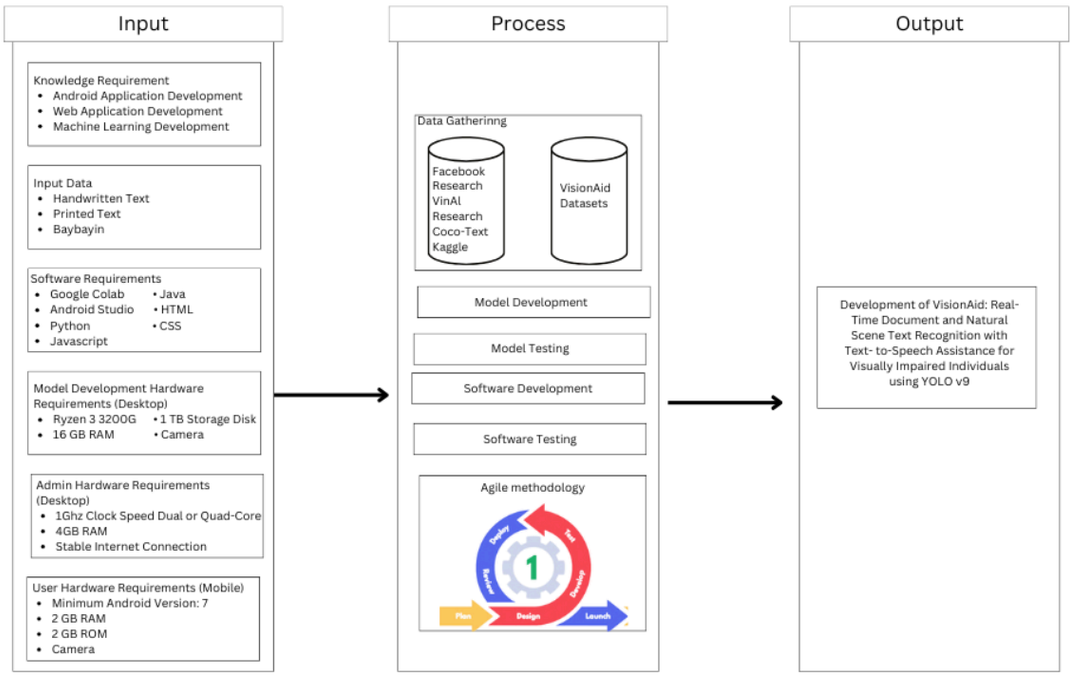


Figure 1. Conceptual Framework of VisionAID

The following Input, Process, and Output of the Program will consist of the following as indicated.

Input

Android application development, Web application, and machine learning development are the required knowledge needed to build this application, input data is handwritten text, printed text, or baybayin, while for dataset will be coming from Facebook Research, VinAI Research, COCO-Text and Kaggle. Software needed for the following study will be as follows: Colab, Android Studio, Python, Java, HTML, CSS, and JavaScript. The hardware required will vary depending on the device and purpose. For development of the model, the minimum requirement is at least Ryzen 3, 3200G with 16 GB of RAM and 1 TB of Storage Disk, while for user and admin use, the minimum requirement of desktop is 1GHz clock speed, dual or quad core processor, 4GB RAM, and a stable internet connection, while for mobile is at least Android 5 and

with a RAM of 2GB and ROM for Mobile. Only the mobile for the end user utilizes a camera for the application to be able to see.

Process

Processing of the following information needs multiple segments as data gathering is needed and once prepared, the researchers will preprocess the data for the model building and would be training the model for the validation, once generated the researchers plan to use the AGILE method for the repetition until certain model evaluation is met.

Output

Result will be the completion of the study which results in the development of the VisionAid: Real-Time Document and Scene Text Recognition with Speech Recognition Assistance for Visually impaired individuals.

Table 1. Benchmarking Analysis

Authors	Title	Problems	Algorithm Used	Findings and Conclusion
Narayan, A., & Raja M. (2021)	*Image Character Recognition using Convolutional Neural Networks*	The objective of this study is to have another method to improve the accuracy of the Optical Character Recognition	Convolutional Neural Network	The returned result had shown great **accuracy of 97.59 which only has a 6.6% loss**. This suggests that the model is best fit and performed well, however for future improvements, it has challenges on text that are not in good condition for the model to recognize.
Aydın, Ö. (2021)	Classification of Documents Extracted from Images with Optical Character Recognition Methods.	This study aims to apply Naive Bayes algorithm for the subsequent recognition for both handwritten and printed documents that are to be extracted.	Naive Bayes	Both the existing and proposed model had achieve high accuracy from character match rates are **approximately 97.4%** however the OCR method had a slightly **higher word rate of 88.52% compare to the MODI counterpart which is the 85.24%** With the text classification of Naive Bayes, the algorithm

				help resulted in approximately 53% accuracy. The study concludes by suggesting various methods to improve accuracy, such as enhancing image quality, cleaning noisy pixels, refining blob detection, increasing the training dataset size, and exploring alternative classification algorithms like Neural Networks.
Gilbey, J. D., & Schönlieb, C.-B. (2021)	An end-to-end Optical Character Recognition approach for ultra-low-resolution printed text images.	Ultra-low-resolution text are text that are completely not ignorable as these things can occur and having the ability for it to be broken down and predicted by Neural	Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN)	The findings of the study showed significant improvements in OCR performance on low-resolution images. **For 60 dpi images, the character error rate was reduced by 64%, and the word error rate was reduced by 73%. For 75 dpi images, the character error rate was reduced by**

		network is something prevalent to the OCR models		**35%, and the word error rate** was reduced by 51%.
Hemanth, G., Jayasree, M., Venii, S., Akshaya, P., & Saranya, R. (2021)	CNN-RNN BASED HANDWRITTEN TEXT RECOGNITION. *ONLINE ICTACT JOURNAL on SOFT COMPUTING*,	Handwritten text recognition (HTR) aims to digitize handwritten and system generated fonts to overcome challenges such as diverse styles, fonts and character distortions.	1. Convolutional Neural Network (CNN), 2. Recurrent Neural Network (RNN) with applied Short-Term Memory (STM) model applied 3. Connectionist Temporal Classification (CTC)	The proposed model had an **accuracy rate of 98% rate for word recognition surpassing existing models.** The study recommends the use of hybrid datasets encompassing diverse writing styles, fonts, and text distortions to further extend the system's capabilities.
Butt, H., Muhammad R. R., Muhamma	Attention-Based CNN-RNN Arabic Text Recognition	To develop a model that is able to recognize	1. Convolutional Neural Network	The Attention-Based CNN-RNN for the Arabic text recognition from natural scenes had a high character

d J. R., Muhammad J. A., & Haris, M. (2021)	from Natural Scene Images.	Arabic text using deep learning algorithms such as CNN and RNN capable of accurately recognizing Arabic text from natural scene images.	(CNN), 2. Recurrent Neural Network (RNN)	recognition rate of 98.73%. The findings suggest effective features as mentioned from the accuracy that can be utilized to learn other sets of text and to achieve high accuracy for doing so.
Pino, R., Mendoza, R., & Sambayan, R.. (2021, February 15).	Optical character recognition system for Baybayin scripts using support vector machines.	This study focuses on the OCR system specifically tailored to understand the Baybayin Scripts, a traditional script used in the Philippines that contains various noises and variations in writing styles.	Support Vector Machine	The proposed SVM model achieved high performance both in binary and multi class classification for baybayin scripts through testings on 1,100 randomly chosen images the accuracy gathered is 98.41% Accuracy. Thus this study suggests a solid text character recognition for the baybayin scripts to be read by OCR.

Fateh, A., Fateh, M., & Abolghasemi V. (2023).	Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection.	This study aims to recognize the regions of text documents and non textual regions within documents while utilizing the OCR system as suggest much more advanced techniques that struggles with accurately identifying text in complex layouts such as newspaper and dictionaries.	1. YOLOv3 2. SSD 3. Faster RCNN 4. Text Line Detection (TLD)	The findings suggest that the integration of deep learning models for region detection and TLD for text preprocessing is a promising approach for enhancing OCR technologies, more information can be found directly on the study and results shown as those that are in bold are greater as this shows lower error rates.
Wei, J., Zhan, H., Tu, X., Lu, Y., & Pal, U. (2023)	*Scene Text Recognition with Image-Text Matching-guided Dictionary*	In documents often aligning visuals with a ground truth are represented, thus having a OCR that is	1.Scene Image-Text Matching (STIM) 2. Image-Text	Application of the SITM network into scene text recognition system has significantly improved the accuracy and robustness compared to existing traditional methods which

		able to detect multiple text without region can lead to low accuracy and increased errors.	Contrastive Learning.	has accuracy of 93.8% which outperforms the 92.1% on six mainstream benchmarks. concluding that the SITM network in aligning visuals and textual features enhances the prediction and accuracy and errors in scene text recognition.
Zhu, W., Sokhanda n, N., Yang, G., Martin, S., & Sathyanar ayana, S. (2022)	DocBed: A Multi-Stage OCR Solution for Documents with Complex Layouts. *Proceedings of the ... AAAI Conference on Artificial Intelligence*,	The following study hopes to develop an OCR solution capable of accurately reading and processing documents with intricate layouts, which the goal is to understand the layout of the following documents that	1. Fully Convolutiona l Network (FCN). 2. Separator Detector (SepDet) 3. Post Processing Methods such as Geometric Approach and Heuristic	As shown within the results of the research each model with algorithms had different result that were best on each region and segments as SETR-Heuristic provide best result for text segmentation other are offer best sequential ordering of text, thus the study concludes that different models have distinct strengths depending on specific application requirements, such as accuracy, inference time, and sensitivity to layout

		are presented to reduce error and improve sequencing of document content	Approach	classification.
Vidhale, B., Khekare, G., Dhule, C., Chandankhede, P., Titarmare, A., & Tayade, M. (2021)	*Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks*.	The objective of this study is to develop an OCR system using CNN to accurately identify and translate handwritten text and printed text documents from various regional language into english, which also includes diverse type of language writing converting them to the	Convolutional Neural Network	The results provide high accuracy from the 10,000 images provided from the MNIST dataset which achieved 99% accuracy on the testing data.

		universally understandable format which is english		
Vilvara, R. A., Hammond , D. S. C., Santos, F. M., & Alar, H. S. (2022)	*Baybayin script word recognition and transliteration using a convolutional neural network. City of Makati, Philippines: University of Makati.*	The surge of interest of people to learn the ancient scripts of Filipinos suggested new needs for character recognition in the field of computers	Convolutional Neural Network,	The model proposed achieved results of an **accuracy of 97.4% for character classification, 96.15% for word detection, and 91.54% for word transliteration.** Showing further improvements may be needed as existing models from VGG16 and Levenshtein Distance had achieved greater research is its own areas of character recognition for VGG16 and Levenshtein distance for word detection.

This study will utilize Optical Character Recognition (OCR) with the application of Convolutional Neural Network (CNN) to achieve remarkable accuracy and build the model from Optical Character Recognition (OCR) as suggested by Aydın, Ö. (2021), which furthering improvement and application towards baybayin script writing and

Filipino text recognition. From experimentation and application of the following studies which had been mentioned above, from studies of Narayan, A., & Raja M, had achieve higher accuracy with the application of CNN and Recurrent Neural Network (RNN), as indicated above with the works of Hemanth, G., Jayasree, M., Venii, S., Akshaya, P., & Saranya, R. (2021) improving their model with the application of CNN and achieving 98% accuracy and even showing high accuracy on the utilization of (RNN). Utilization of CNN is also applicable to numerical character recognition as suggested by Vidhale, B., Khekare, G., Dhule, C., Chandankhede, P., Titarmare, A., & Tayade, M. acquiring 99% accuracy on the provided images from MNIST datasets. With these following studies mentioned above, had created a robust foundation for the algorithms to be applied on the model which the researchers had proposed. Furthermore, comprehensive research needs to be done for the application of OCR methods which applies intriguing insights about models that had a text line correction from Fateh, A., Fateh, M., & Abolghasemi V (2023), applied. The results that are shown the result which provides the error rate on the given model against the existing model, finally show great improvement on the previous model of the Optical Character Recognition (OCR).

From the study of Gilbey, J. D., & Schönlieb, C.-B (2021). , which delves into ultra-low-resolution printed text images as this paper resize the images, and have a model that is able to read low-resolution images with the application Gaussian filtering and use of standard deviation for blurring and upscaling, this shown great improvement and application, especially if there are printed text documents which possibly hard to read especially for visually challenged individuals. The application of Baybayin Script recognition with individually recognizing each character from Pino, R., Mendoza, R., & Sambayan, R., and further showing great recognition as Butt, H., Muhammad R. R., Muhammad J. R., Muhammad J. A., & Haris, M. applied the same recognition that uses the same algorithm which is CNN and RNN achieving accuracy from Arabic text of 98.73% and 98.41% on Baybayin scripts.

As scenery recognition is something the researchers acknowledge, a clear understanding of environment and recognition of text from murals, signs and everything else which text are written on a non document, from Wei, J., Zhan, H., Tu, X., Lu, Y., & Pal, U as shown that achieved the accuracy of 98.8% accuracy, the researchers will try to implement Scene Image-Text Matching (STIM) and Image-Text Contrastive Learning. for further understanding of text on difficult background and possibly lighting. Which leads to the focus of the study of the implementation of You Only Look Once (YOLO) on the model as introduced by Zhu, W., Sokhandan, N., Yang, G., Martin, S., & Sathyanarayana, on their study which applies the character recognition and line segmentation on text documents for better understanding the text and sequence of the following texts.

Synthesis

The developmental of the new technologies nowadays cannot be ignored, as new improving technological advancement are always being offered to the field of research such as the application of Optical Character Recognition (OCR) which is proven reliable by Aydın, Ö. (2021) on this study, recommended that to use neural network on to improve the Optical Character Recognition and and this technology has emerged as a tool in making textual information accessible to visually challenged individuals. Following recommendations from Narayan, A., & Raja M. (2021), the utilization of Convolutional Neural Network to be applied on OCR for better result , however, was also advised to use better segmentation on the research of Butt, Muhammad R. R., Muhammad J. R., Muhammad J. A., and Haris (2021), as the existing system can already perform very well despite of their findings of the utilization of attention mechanism enhanced RNN, it is much more better to focus on harsh environment text recognition as already seen on the study of Gilbey, J. D., & Schönlieb, C.-B. (2021), a better segmentation would be better for application of Optical Character Recognition (OCR). While Zhu, Sokhandan, Yang, Martin, and Sathyanarayana (2022) already proposed the utilization of Document Layout Analysis (DLA) and Text Line

Detection (TLD) for better understanding of printed out documents and actual sceneries that can be harsh for the model to understand. It is already suggested by Fateh, A., Fateh, M., & Abolghasemi V. (2023). that the utilization of the following segmentation can provide excellent performance based on the benchmark shown on the table 1, while this study proves on the application of the Document Layout Analysis (DLA) and Text Line Detection (TLD), it was also recommended by the researchers to use on bigger data sets and use it on a different language as well which is proven well by Vilvara, R. A., Hammond, D. S. C., Santos, F. M., & Alar, H. S. (2022), and Pino, Mendoza, and Sambayan (2021) with the help of Support Vector Machine. Finally the researchers aims to create a model that will utilize YOLO v9 which is for object detection to apply machine learning algorithms with a much more bigger datasets gathered from Vinai Datasets and Publaynet.

1.Image Character Recognition using Convolutional Neural Networks. In this study the application of CNN returns high accuracy finding a good foundation of an algorithm that this study will use.

2. **Classification of Documents Extracted from Images with Optical Character Recognition Methods.** Both existing and proposed model had character match rate which are approximately the same under 97.4% however the proposed OCR method had slightly higher rate of 88.52% compared to the MODI counterpart and with the help of the algorithm of Naive Bayes, it improved the accuracy gain 53% accuracy which can be utilized in the study in the future

3.**An end-to-end Optical Character Recognition approach for ultra-low-resolution printed text images.** With the application of the said algorithm within the study, findings had shown significant improvement of OCR performance on low resolution images which can contribute to the model proposed by the group greatly especially for blurred images.

4. **CNN-RNN BASED HANDWRITTEN TEXT RECOGNITION. ONLINE ICTACT JOURNAL on SOFT COMPUTING**. The proposed model had an accuracy of 98% which is extremely high for the word recognition rate of existing models, which would greatly improve character recognition on the proposed model.

5. **Attention-Based CNN-RNN Arabic Text Recognition from Natural Scene Images.** As the proposed model helps the utilization of CNN and RNN based for character recognition on sceneries, which are from signs and outside this can contribute to image recognition which are outside of handwritten text and printed documents that can be improved.

6. **Optical character recognition system for Baybayin scripts using support vector machines**. As this study contributes to the proposed model of the researchers which is to have the model understand Baybayin scripts which is a Filipino manuscript that people can also give attention to as a feature on the program.

7. Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection. As text alignment is very complicated on dictionaries and newspapers, even on other documents may have complicated text and ordering which can give errors for the Text to speech factor of the program, this opens new opportunities which the program is open to create and acknowledge.

8. **Scene Text Recognition with Image-Text Matching-guided Dictionary.** This helps to improve scene alignment which can contribute to scene image text matching that has given very high accuracy of 93.8 % which outperforms existing models, this study can contribute to the utilization of high accuracy of text extraction from scenes that may be challenging to capture or extract text.

9. **DocBed: A Multi-Stage OCR Solution for Documents with Complex Layouts.** Having the availability of the text available for the user is a feature the model should also have as if the user prefers to read it on a higher image quality or enhanced text

instead of hearing it, the extraction of text with the utilization of CNN can help to enable the user to read the text on a much better scene.

10**. Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks.** The basics of number writing can be complicated especially if there are certain ways how a certain ethnic and culture has their own way to write their number systems, this can greatly contribute to the proposed model as this returns high accuracy of 99% on the testing data out of 10,000 Images which the researchers had used from the databases available.

11. **Baybayin Script word Recognition and Transliteration Using a Convolutional Neural Network**

From the paper of Vilvara, R. A., Hammond, D. S. C., Santos, F. M., & Alar, H. S. (2022), Convolutional Neural Network (CNN)-based transliteration model designed to convert Baybayin script into Filipino words. The model was trained on a dataset specific to Baybayin character classification, word detection, and transliteration, and evaluated in these categories. The results show that the proposed model achieved an accuracy of 97.4% in character classification which can be utilized and applied since the proposed model hopes to implement baybayin character recognition.

Chapter 3

DESIGN AND METHODOLOGY

This chapter presents the methodology of the study. This part elaborates the different approaches of methods to prove the efficiency of the algorithm. Contents in this chapter states on how the researchers' approach in terms of research and development to the study.

Research Design

This study will follow the mixed-method research design, integrating both experimental and descriptive research design to develop the model by incorporating the chosen algorithm, and evaluate the application.

According to the definition given by Saigo(2023) of experimental research design, a research approach used to investigate the interaction of dependent and independent variables, which can be utilized to determine a cause and effect relationship. This method allows the researchers to control and manipulate variables to observe their effects. In this study, the researchers modified and customized the YOLO v9 object detection model by incorporating Document Layout Analysis at the initial stage of building the OCR model. By integrating layout analysis, the proposed system is better equipped to identify and differentiate between various elements of a document, whether it is handwritten or printed. The proponents further adapted YOLO v9 to recognize and classify components like text blocks, and tables, which improves the model's ability to understand document structure.

While descriptive research, as defined by Hassan(2024), it aims to describe or record the traits, behavior, attitudes, opinions, or perceptions of a group or population under study, which is useful for VisionAID to gather insights regarding the visually impaired individuals who use the application.

Research Methodology

Figure 2. Agile Software Development (Petit, 2021)

Figure 2 illustrates the developmental stages that the researchers must go through in order to build VisionAID.

Plan

During this phase, the researchers gather knowledge defining the objective and scope of the application. This involves identifying key features such as handwritten and printed document recognition, identification of document layout for text blocks, and tables. The researchers also select the appropriate algorithm in order to build the application and model, this includes the YOLO v9 model for a custom OCR and other necessary libraries, as well as the datasets needed in order to build the proposed model.

Design

In this phase, the application's user interface and experience are conceptualized through wireframes and prototypes. This helps to outline the application's layout, navigation flow, and interactive elements, ensuring it is user friendly and accessible to visually impaired users. The researchers also define the system architecture, planning

how different components such as the model, user interface, and backend servers will integrate and communicate.

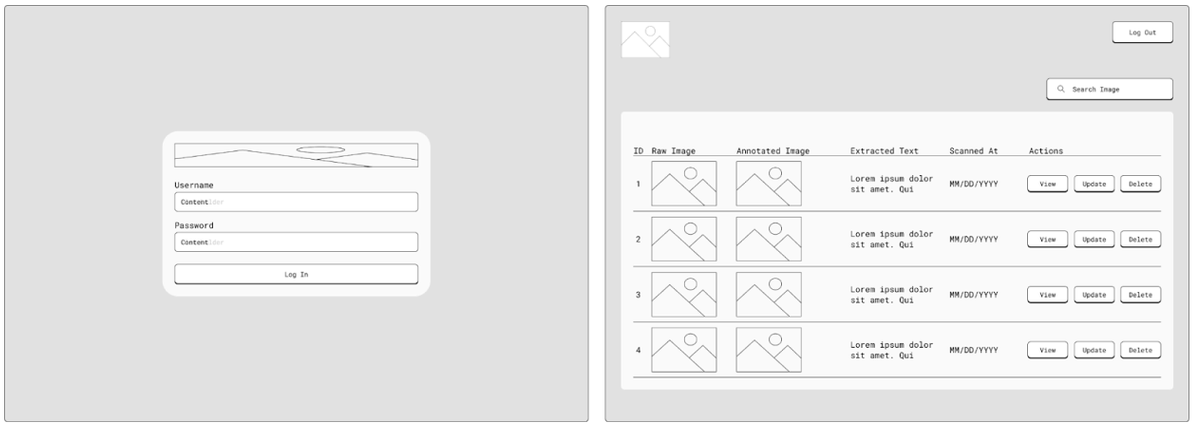


Figure 3. Low-Fidelity Wireframe of VisionAID Admin Web Application

Shown in Figure 3. is the Low-Fidelity Wireframe of the VisionAID Admin Web Application, it consists of two screens, the login screen, and the dashboard screen. The login screen, located on the left, provides a simple interface where administrators of the application can enter their credentials and log in to the application. Once logged in, administrators are redirected to the Dashboard screen on the right, which serves as the central hub for managing scanned image data. Administrators can use the dashboard in order to keep track of the scanned images, they have control whether they would like to view/update/delete a record in the database, these scanned images can be used to improve the model by training it with new data.

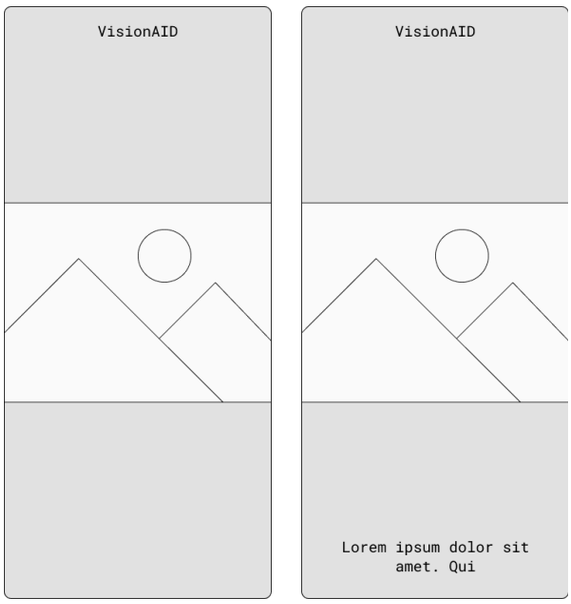


Figure 4. Low-Fidelity Wireframe of VisionAID User Mobile Application

Figure 4. shows the low-fidelity wireframe of VisionAID application designed for the client or end user. The left panel displays a screen where the user can utilize their camera to scan an image. Meanwhile, the right panel shows the text that has been scanned or extracted from the image, which is then read back to the user.

Algorithm Design and Model Architecture

In this study, the researchers will be utilizing a combination of deep learning algorithms. The following algorithms are as follows:

Convolutional Neural Network (YOLO Architecture): This machine learning algorithm allows VisionAID to identify and localize objects within an image or video stream, specifically the *YOLO v9* Architecture will be utilized by VisionAID, to consider the computational performance of this architecture in order to give real time detection. According to Wang, Yeh, and Liao (2024), compared to its predecessor, YOLOv9 is a key step forward in real-time object detection, bringing notable enhancements in efficiency, precision, and flexibility.

Optical Character Recognition(YOLOv9 with Document Layout Analysis): This algorithm enables VisionAID to recognize, analyze, and extract text from an image allowing users to read, and navigate text prompted on the application for better information extraction.

Data Gathering Procedure

Datasets from Publaynet, COCO-TEXT, VinAI's Research, Facebook Research, and Kaggle will be collected and merged into one whole new dataset, these data will contain the handwritten text, document text, natural scene text, and a combination of the mentioned type of text. This dataset will be used to train and evaluate the VisionAID's performance on tasks like object detection, text recognition, and document analysis.

Instruments and techniques

Data will be collected using a combination of quantitative and qualitative methods. The following instruments and techniques are as follows:

Datasets

Pre-existing datasets containing text, and images will be used for building the machine learning model of Vision AID.

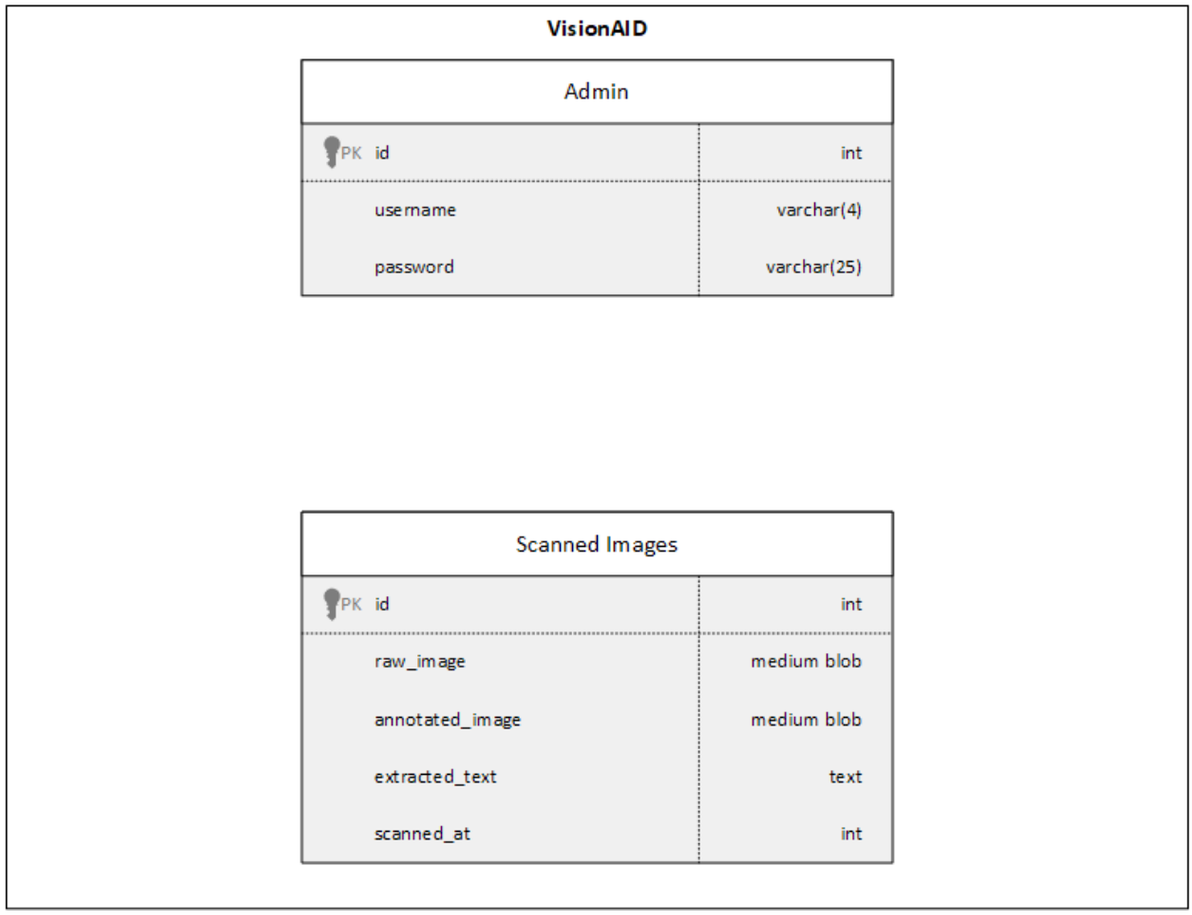


Figure 5. VisionAID Database Structure

Figure 5. shows the database structure of the VisionAID application, where it has two main entities, Admin, and Scanned Images. The admin entity stores the information about the administrator of the application, it consists of an id field as the primary key to uniquely identify the user, username field with maximum length of 4 characters, password with maximum length of 25 characters to securely store the credentials. The scanned image is a collection of images scanned by the user using the VisionAID application, the mobile application automatically sends the scanned image from the mobile application to the backend server for storing the image data. It includes id for uniquely identifying the image data, raw_image for storing the raw unextracted

and unannotated image, annotated_image for the visual representation of what the model sees, and extracted_text for text extracted from the annotated image of the model, and scanned_at for storing the date and time of when the image was scanned at.

Figure 6. VisionAID Data Flow Diagram Level 0

Figure 6. shows the overview of the system through a Data Flow Diagram Level 0. It shows the interaction between the user and the model through the VisionAID system, where the initial input starts from the user scanning images and the system preprocessing it before analyzing using the model for extraction of text and analyzation of document layout, which will then be read back to the user using text to speech.

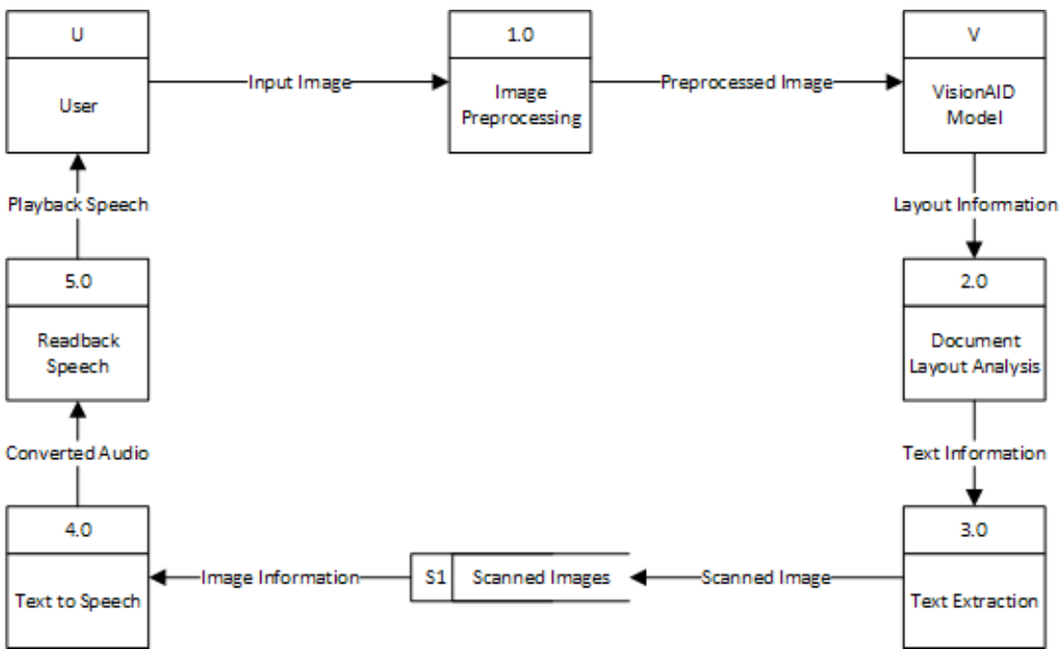


Figure 7. VisionAID Data Flow Diagram Level 1

Figure 7. shows a more detailed view of the VisionAID system, compared to DFD 0, it shows the breakdown of the processing into smaller pieces. Similarly, it starts with the image as an input which will then be processed, and analyzed using the VisionAID system. The scanned image will then be saved into the database for future reference, as well as the information of the scanned image will be converted into an audio using text to speech which will be read back to the user.

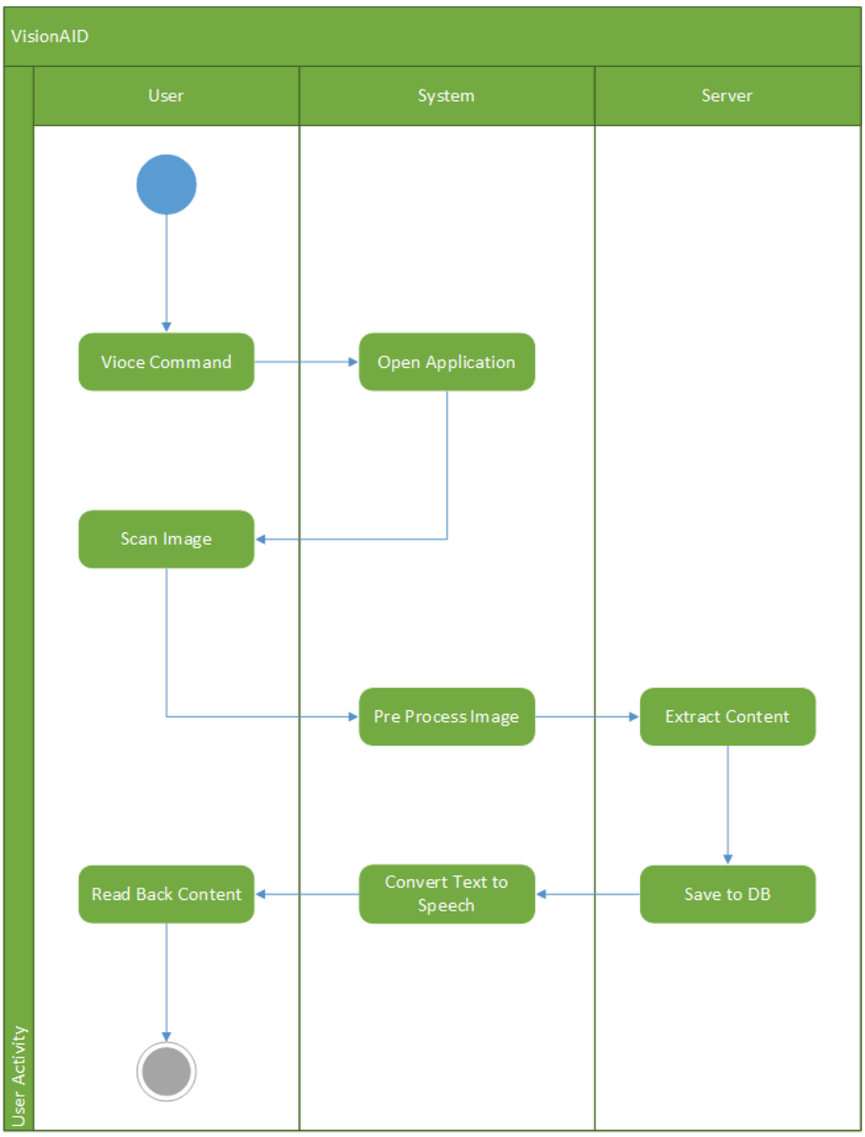


Figure 8. User Activity Diagram

In this Figure 8. shows the sequence of interaction between the end-user, VisionAID application or system, and the Server. Initially the user sends out a voice command in order to open the application, which the system process and opens the VisionAID application ready to scan an image, after the image has been scanned, it will be pre-processed before sending to the server for the utilization of the VisionAID model, and extraction of text and logging of the image in the database. Lastly, the model will then send back the extracted text from the image to the application which will then read back its content to the user.

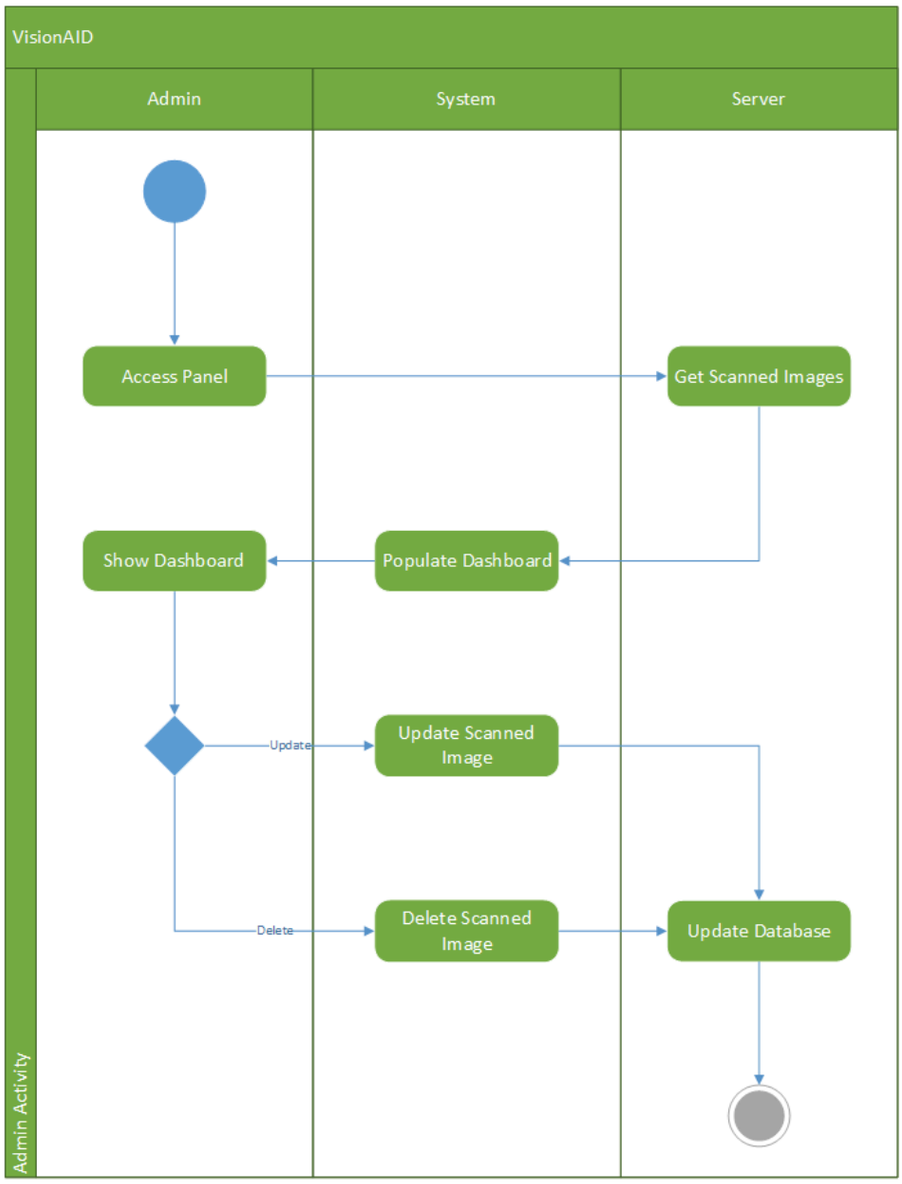


Figure 9. Admin Activity Diagram

An admin activity is shown in Figure 9. starting with accessing of the admin panel, and the VisionAID server will get all data in the database for the scanned images, which will then populate the html layout for the dashboard that will be shown or displayed to the user. The administrator has a choice of whether they will delete or update data which then be processed accordingly and will be updated in the database.

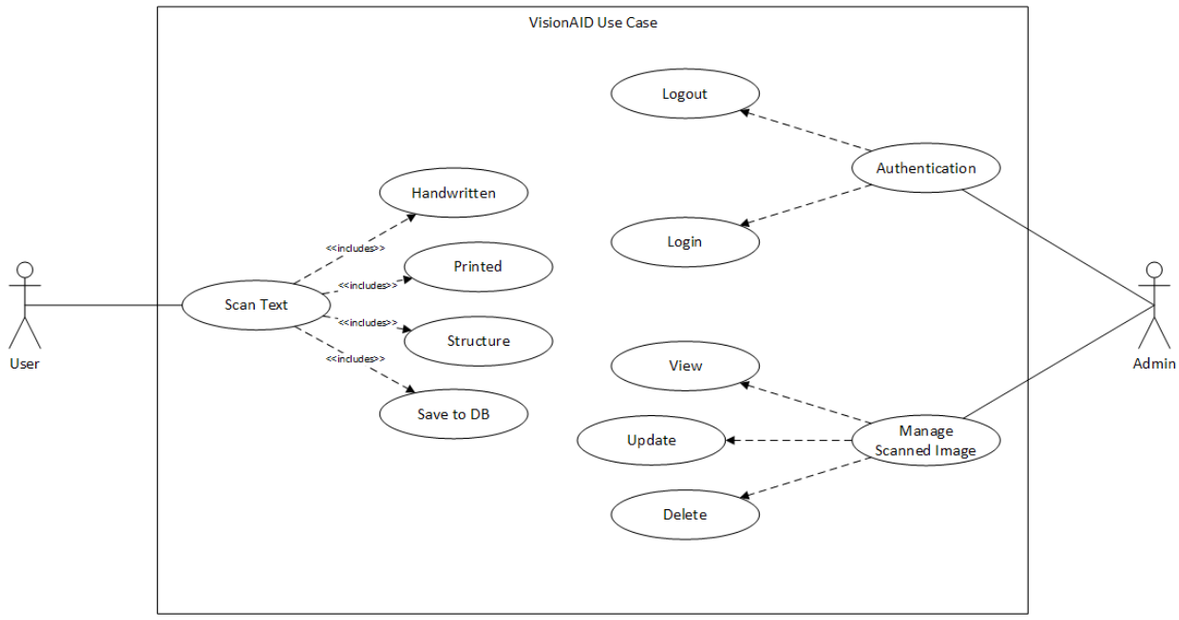


Figure 10. VisionAID Use Case Diagram

Figure 10. depicts a use case diagram that visualizes how different actors interact with the VisionAID application, highlighting key features of the application.

Actors:

- **User** - Visually impaired individuals who use the VisionAID application.
- **Admin** - Overall or system admin of the VisionAID application.

Use Cases:

- **Scan Text**
    - **Handwritten** - Allows the user to scan handwritten documents.
    - **Printed** - Allows the user to scan printed documents.
    - **Structure** - Analyze the structure of the document.

    - **Save to DB** - Allows the user to save the scanned image to the system database.
- **Authentication**
    - **Login** - Admin can enter their credentials to access the system dashboard.
    - **Logout** - Exit out of the application session.
- **Manage Scanned Images**
    - **View** - View all of the scanned image/s from the database.
    - **Update** - Updates the attribute of the scanned image/s from the database.
    - **Delete** - Delete scanned image/s from the database.

Relationships:

- **User** - Has access to the core functionalities of the VisionAID application, allowing them to scan a document and utilize the model to extract and analyze the document structure.
- **Admin** - Has a higher privilege of accessing the application, allowing view/update/delete of data from the database.

Develop

Model and application development is done during this phase. The development will be done to the researcher's specifications.

The model will be developed using the python programming language, and its datasets accordingly. This includes data pre-processing, model training, and model testing.

The mobile application will be built using android's native language which is Java, while for the web application, which is for admin to use, HTML, CSS, and JavaScript will be used. This includes the integration of frontend and backend connection, ensuring smooth communication between both ends.

For the backend server, the application will be built using Python Flask, a lightweight web framework designed to be run on a WSGI(Web Server Gateway Interface) server. The backend will operate on an Ubuntu operating system.

The WSGI server that will be utilized in this research is Gunicorn, a python WSGI server for UNIX. which will manage the Flask application's execution and handle HTTP requests.

NGINX will also be utilized as the web server and reverse proxy, on top of Gunicorn, it will handle incoming HTTP traffic, server static files, and route dynamic requests to Gunicorn.

All of these components are connected, ensuring the efficiency of handling web requests and scalability of the applications.

Test

During this phase, the researchers will systematically evaluate VisionAID across multiple test cases. This comprehensive testing process is designed to thoroughly assess and ensure the reliability and quality of the software or the application itself.

Evaluation Procedure

In evaluating the performance of VisionAID's machine learning model and overall performance, the following metrics will be used:

Accuracy

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives}$$

Equation 1. Accuracy

Accuracy measures all correct predictions the model makes. It is defined as the total number of correct predictions divided by the total number of instances.

Precision

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Equation 2. Precision

Precision measures the accuracy of the positive predictions made by the model. It is defined as the number of true positives divided by total predicted positives.

Recall

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Equation 3. Recall

Recall, also known as *sensitivity,* measures the ability of the model to identify all relevant instances. It is defined as the number of true positives divided by the actual positive.

F1 Score

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Equation 4. F1 Score

F1 Score is defined as the harmonic average between precision and recall.

Deploy

In this phase, the focus is on getting the application up and running in accordance to the specified requirements. It involves checking that all of the intended features are implemented and functioning as expected.

Review

Lastly, during this phase, the researchers gather feedback from the visually impaired individuals on what is needed to be improved on the system. Also during this phase, the researchers also assess all of the implemented features, and fix any bugs or errors in the system if there's any.

User Satisfaction of the application will be measured using Likert Scale, this involves users rating their satisfaction on a scale, where a higher number indicates greater satisfaction, shown in Table 1.

Table 1. User Satisfaction Score

Score	Interpretation
1	Strongly Dissatisfied
2	Dissatisfied
3	Neutral
4	Satisfied
5	Strongly Satisfied

The average score will be used to quantify the overall user satisfaction of VisionAID as shown in Table 2.

Table 2. Overall User Satisfaction

Range of Mean Overall Score	Interpretation
1.50 and below	Strongly Dissatisfied
1.51 - 2.50	Dissatisfied
2.51 - 3.00	Neutral
3.51 - 4.00	Satisfied
4.51 - 5.00	Strongly Satisfied

ISO 25010

A standard for software product quality that defines a model for evaluating software. It includes eight quality characteristics such as functionality, reliability, usability, and performance efficiency, to ensure comprehensive assessment of software quality.

References:

Aydin, O. (2021). Classification of Documents Extracted from Images with Optical Character Recognition Methods. ArXiv.org. https://arxiv.org/abs/2106.11125

Burton, M. J., Ramke, J., Marques, A. P., Bourne, R. R. A., Congdon, N., Jones, I., Ah Tong, B. A. M., Arunga, S., Bachani, D., Bascaran, C., Bastawrous, A., Blanchet, K., Braithwaite, T., Buchan, J. C., Cairns, J., Cama, A., Chagunda, M., Chuluunkhuu, C., Cooper, A., ... Faal, H. B. (2021). The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. The Lancet Global Health, 9(4), e489–e551. https://doi.org/10.1016/s2214-109x(20)30488-5

Butt, H., Muhammad R. R., Muhammad J. R., Muhammad J. A., & Haris, M. (2021). Attention-Based CNN-RNN Arabic Text Recognition from Natural Scene Images. Forecasting, 3(3), 520–540. https://doi.org/10.3390/forecast3030033

Cleveland Clinic. (2020). *20/20 Vision: What It Means & Corrective Methods*. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/8561-2020-vision

Fateh, A., Fateh, M., & Abolghasemi V. (2023). Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection. Engineering Reports. https://doi.org/10.1002/eng2.12832

Gilbey, J. D., & Schönlieb, C.-B. (2021). An end-to-end Optical Character Recognition approach for ultra-low-resolution printed text images. ArXiv.org. https://arxiv.org/abs/2105.04515

Hassan, M. (2024). Descriptive Research Design – Types, Methods and Examples. Research Method. https://researchmethod.net/descriptive-research-design/

Hemanth, G., Jayasree, M., Venii, S., Akshaya, P., & Saranya, R. (2021). CNN-RNN BASED HANDWRITTEN TEXT RECOGNITION. ONLINE ICTACT JOURNAL on SOFT COMPUTING, 1. https://doi.org/10.21917/ijsc.2021.0351

Mcleod, S. (2023). Likert Scale Questionnaire: Examples & Analysis. https://www.simplypsychology.org/likert-scale.html

Narayan, A., & Raja M. (2021). Image Character Recognition using Convolutional Neural Networks. https://doi.org/10.1109/icbsii51839.2021.9445136

Petit, C. (2021). Agile Software Development: What It Is and the Benefits It Offers. Softrizon. https://www.softrizon.com/blog/agile-software-development/

Philippine Statistics Authority. (2023, December 22). Preliminary 2023 First Semester Official Poverty Statistics. https://www.psa.gov.ph/statistics/poverty

Pino, R., Mendoza, R., & Sambayan, R.. (2021, February 15). Optical character recognition system for Baybayin scripts using support vector machine. ResearchGate; PeerJ. https://www.researchgate.net/publication/349323182_Optical_character_recognition_system_for_Baybayin_scripts_using_support_vector_machine

ROQUE Eye Clinic. (2024). Persons with disabilities. https://www.eye.com.ph/about-us/policies/pwd/

Saigo, H. (2023). Experimental Research Design | Definition, Components & Examples. Study.com. https://study.com/academy/lesson/the-true-experimental-research-design.html

STATISTA. (2024, January). Philippines: major mobile OS by market share 2024. https://www.statista.com/statistics/931129/philippines-mobile-os-share

Tesseract-OCR. (2024). Tesseract. GitHub. https://github.com/tesseract-ocr/tesseract

Tigerschiold, T. (2022). What is Accuracy, Precision, Recall and F1 Score? Labelf.ai. https://www.labelf.ai/blog/what-is-accuracy-precision-recall-and-f1-score

Vidhale, B., Khekare, G., Dhule, C., Chandankhede, P., Titarmare, A., & Tayade, M. (2021). Multilingual Text & Handwritten Digit Recognition and Conversion of

Regional languages into Universal Language Using Neural Networks. https://doi.org/10.1109/i2ct51068.2021.9418106

Villanueva, M. A. (2022). Lost focus in the fight vs blindness. Philstar.com; Philstar.com. https://www.philstar.com/opinion/2022/11/30/2227378/lost-focus-fight-vs-blindness

Vilvar, R. C., Hammond, D. S. C., Santos, F. M. R., Alar, H. S. (2022). Baybayin Script Word Recognition and Transliteration Using a Convolutional Neural Network. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4004853

Wang, C.-Y., Yeh, I-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. https://arxiv.org/pdf/2402.13616

Wei, J., Zhan, H., Tu, X., Lu, Y., & Pal, U. (2023). Scene Text Recognition with Image-Text Matching-guided Dictionary. ArXiv.org. https://arxiv.org/abs/2305.04524

WHO. (2019). World report on vision. https://www.who.int/docs/default-source/documents/publications/world-vision-report-accessible.pdf

Zhu, W., Sokhandan, N., Yang, G., Martin, S., & Sathyanarayana, S. (2022). DocBed: A Multi-Stage OCR Solution for Documents with Complex Layouts. ArXiv.org. https://arxiv.org/abs/2202.01414

Vilvara, R. A., Hammond, D. S. C., Santos, F. M., & Alar, H. S. (2022). *Baybayin script word recognition and transliteration using a convolutional neural network*. City of Makati, Philippines: University of Makati. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4004853

DATASETS

VINAI Dataset: from https://github.com/VinAIResearch/dict-guided?tab=readme-ov-file

PubLayNet Dataset: https://github.com/ibm-aur-nlp/PubLayNet