

## **DIGITAL DATA CLASSIFICATION AND EXTRACTION FOR RECORDS MANAGEMENT OF PAPS AND CHILD DOCUMENTS USING TESSERACT OCR**

Jennifer E. Sabugaa (1), Glenn S. Lahayon (2)

<sup>1</sup> Faculty, Caraga State University, Ampayon, Butuan City

<sup>2</sup> Student, Caraga State University, Ampayon, Butuan City

Email: [jenniferespanola@gmail.com](mailto:jenniferespanola@gmail.com), [glennlahayon@gmail.com](mailto:glennlahayon@gmail.com)

**KEY WORDS:** Data Classification, Data Extraction, Case Visualization, Optical Character Recognition, Notification

**ABSTRACT:** The Philippine government offices are mandated to make business easier through the Republic Act No. 11032 or the “Ease of Doing Business and Efficient Government Service Delivery Act of 2018”, signed by President Rodrigo Roa Duterte on May 28, 2018. This new law is an amendment of Republic Act 9485 also known as “The Anti-Red Tape Act of 2007”. One of the most important features of this law is the standardized deadline for government transactions [H. Fuentes, 2018]. It has a big impact particular in Alternative Parental Care (APC) handled by the Department of Social Welfare and Development (DSWD) Caraga Field Office, who will be responsible for child care and placement services provider to children with difficult circumstances. To comply with the Anti-Red Tape Act, some adoption processes were improved with the integration of the emerging technologies. Adoption timeline is a sequence of adoption that is typically visualized by plotting them along an axis at the instant of interval (Nguyen, Xu, Wlaker, Wong, 2015). Timeline visualization will be good if match with notification to ensure the case handler about the case via email and push notification. The Short Message Administration (SMA) innovation is a standout amongst the most steady and most broadly utilized portable specialized techniques after telephone calls (Olaleye, Olaniyan, Eboda, & Awolere, 2013). While

the email notification focusses on retention, increasing the lifetime value of a customer and serves as the best marketing tool (Yu, 2014). Included in the improvement was the extraction of data from a pdf document using Tesseract Optical Character Recognition vs. Convolutional Neural Network (CNN) so that the user-in-charge need not to manually input the APC documents. It will also deploy the server-based digitized records management for easy accessibility and providing secure storage of the digitized documents. A testing on the reliability of the extraction result was made and based on the result given the resolution with 150 dots per inch, it returned an accuracy rate of 99.23619895. The accuracy rate of 150 DPI resolution is higher than of 300 DPI and even on a clean document generated from word applications.

### **INTRODUCTION**

In the digital word, knowledge is divided into two parts the binary strings of 0s and 1s, which make up the genetic code of digital data which people, created, manipulated and shared (Kannappanavar & Rajanikanta, 2010). The consumption of data is essential than its longevity. Considering the fact that every day a volume of data was created using various formats and it is difficult to catch up with a flood of information without the help of computer technology (Thiyam Satyabati Devi, 2005). The right information to the right user at the right time is the aim of professional in recent developments of information and communications technologies (Saima Khan, 2015). The

process of creating digital objects from physical, analog originals by means of a scanner, camera, or other electronic device is called digitization. The assets of digital information are its flexibility and it's very easy to produce an identical copy because it does not decay by virtue of copying (Lingbanan, Importance of Digitization and Content Management, 2016).

Digitized materials contain confidential data that needs to be safe and secured. There are two types of storage that is to be considered upon storing digitized data, which is the cloud storage and local storage. In terms of security and safety, local storage will be considered as an advantage. Upload and download speed also must be considered in storing data. However, the maintenance of local storage expensive (Hedge, 2016). But for an institution like the Department of Social Welfare and Development, it is more important to keep the files safe and secure and a server was already established with the use of Regional Information Communication and Technology Management Unit (RICTMU). According to the study of (Prabhaka & Rani, 2018), institutional repositories are in need due to technological change, significant increase in the overall volume of research, increasing need for archival and access to unpublished information, increase demand to access knowledge objects from anywhere at any time and increasing uncertainty over who will handle the preservation archiving of materials.

In today's era, government offices are mandated to make business easier. President Rodrigo Roa Duterte signed into law the Republic Act No. 11032 or the "Ease of Doing Business and Efficient Government Service Delivery Act of 2018". This new law is an amendment of Republic Act 9485 also known as "The Anti-Red Tape Act of 2007". One of the most important features of this law is the standardized deadline for

government transactions [H. Fuentes, 2018]. The new law has a big impact particular in Alternative Parental Care cases handled by DSWD Field Office Caraga. The adoption timeline must be strengthened in order to comply with the new law. Timeline is a sequence of events that is typically visualized by plotting them along an axis at the instant of interval (Nguyen, Xu, Wlaker, Wong, 2015). Timeline visualization will be good if match with notification to ensure the case handler about the case via email and push notification. The short message administration (SMS) innovation is a standout amongst the most steady and most broadly utilized portable specialized techniques after telephone calls (Olaleye, Olaniyan, Eboda, & Awolere, 2013). While the email notification focusses on retention, increasing the lifetime value of a customer and serves as the best marketing tool (Yu, 2014).

Considering the fact that manual inputs by the user causes a time delay and affects the productivity of the user in performing another task. The adoption case manager is very busy with paper works and deliverables. Time management and using new technology to make things faster is important for them.

For this reason the researcher have designed and developed a system that will get data from the digitized file and save it to the database for fast and easy analyzation. Specifically, a system that would extract and classify data from Alternative Parental Care printed documents; a server-based Alternative Parental Care digitized records management system; a comprehensive visualization on successful APC for top management decision making; a system that will notify the case manager about their case handled and evaluate the system generated output, compare to manual inputs.

## RELATED LITERATURES

Form case managers who personally interview the PAPs and child, the data they gathered will be recorded in the form of the Home Study Report and Child Study Report. This report will be submitted to the Regional office so that the case will be accounted for. Once the case will be verified, it will be saved into the caseload inventory. The caseload inventory will serve as the database of all case that the region is holding.



Figure 1 Current Flow of Adoption Cases

A case study of Optical Character Recognition by Open Source OCR Tool Tesseract was conducted by Chirag Patel, Atul Patel, Ph.D. and Dharmendra Patel of Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA) Charotar University of Science and Technology (CHARUSAT). This study compares tesseract with other commercial Optical Character Recognition (OCR) application, considering the plate number as a parameter of text extraction. It is stated that OCR is converting the printed document into editable text for further processing. The technology was designed to allow the machine to recognize text automatically. Accordingly, Tesseract was one of the open source OCR engine. The process flow of Tesseract OCR starts from the input image and pass to Adoptive Threshold which converts input image into a binary image. Then, the associated segment examination which is utilized to remove character plots. The next thing is to find lines and words which finds character outlines and organize into words.

The processing of image and how it traverses into the engine with a final result of extracted text from an image. This architecture tested with the use of windows command prompt with images that contain English text. Although Tesseract OCR supports various

language, to test the accuracy of OCR the researcher uses the image with the English language.

Another study entitled “Deep Learning classification in Low-Resolution LWIR Imagery via Transfer Learning” was conducted by Abbott R., Del Rincon J.M., Connor B., and Robertson N. of ECIT, Queen’s University Belfast compares the YOLOv2 and Keras CNN. The two systems were evaluated using two datasets with a small dataset of thermal images in low resolution and a large dataset of thermal images in high resolution. YOLOv2 has a precision rate of 52% which is a good start for different datasets. After fine tuning, YOLOv2 increase its precision up to 84%. On the other hand, the Keras initially gives 26% accuracy rate and after fine-tuning, the precision rate raised up to 87%.

The study of Siva Kumar a/l Subramaniam, Siva Kumar a/l Subramaniam, Siti Huzaimah binti Husin, Yusmarnita binti Yusop, Abdul Hamid bin Hamidon of Universiti Teknikal Malaysia Melaka entitled “Real-Time Mailbox Alert System via SMS or Email” revolutionize the use of short messaging technology and electronic mail. Upon receiving a mail, a sensor from a mailbox will communicate to GSM Modem then the GSM Modem will automatically send a notification to the owner of the mailbox via SMS or email. The system uses Zelio Logic Smart Relay (ZLSR) which is a programmable logic controller and designed for a small automated system.

## METHODS

The researcher created a conceptual framework of how the system works. The system uses pdf file which serves as input passes through an OCR engine which extracts text from the document. With extracted text, the significant data will be saved to the database and the document will be uploaded to the server. With data from the database, the case notification and visualization will be generated.

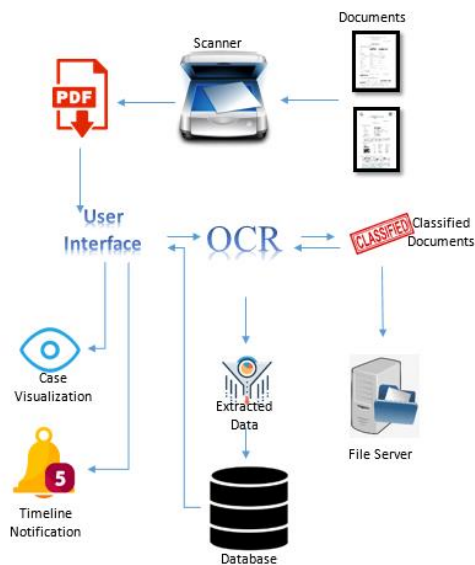


Figure 2. Conceptual Framework

The researcher finds two ways in order to extract data from the document. First is with the use of Convolutional Neural Network (CNN) while the other way is the Optical Character Recognition (OCR). CNN recognizes the text or digit by treating the document as an image and converting the image into a 2-dimensional array. CNN will create a 2x2 or 3x3 matrix and it will traverse to the whole image and the output will pass to a Neural Network to predict the text. While OCR extracts the text from an image by converting the whole image into binary pixels and from that binary pixels and compare the recognized sets to its library to predict the text and set of words.

In order to extract data from the APC document, the system will allow the user to upload an APC document and it will be placed in a canvas. From the canvass, the user can view to what page of the document will be extracted. From the selected document, the system will transform a pdf page into an image. Then with the use of Tesseract with optical character recognition, the text will be extracted from the data from the document. The extracted data will then be saved to a database, while the pdf file will be uploaded to the server for safe keeping and possible use in the future.

Optical Character Recognition is the technology used to convert text from images into a machine-encoded format. The flow

starts from an input image is being processed converting the original image into a binary image and put into an OCR. After OCR processing the text from the image will be extracted. In order for the OCR engine to support a wide range variety of image format, the leptonica library was used. This library is an open source library used for image processing applications. Feature detection is one of OCR approach which analyzing the lines and strips of the character. Another approach is the pattern recognition which analysis of the lines of text and converts the image into a binary matrix where white pixels are 0's and black pixels are 1's.

Document classification is the prediction of what type of document is being processed in the system. The documents need to be converted as an image in order to train the model and to predict the type of image that is currently processed. To train a model the researcher converts the documents into a group of images. Form scanned documents in the form of pdf, the selected page of a document was saved as a jpg file with the use of photoshop. After creating a group of images, the researcher uses keras in order to train the model with tensorflow on the backend. Keras works by defining a sequence of layers in a network by creating a sequential class and adding new layers. This will be done by creating an array of layers and pasting it to the constructor of the sequential model. The first layer of the system must characterize the number of inputs to expect, the way that it is specified defer on the type of network. The layer can also be added with an activation function. This layer will define the format of prediction. After defining the network, it will be compiled which means it will be transformed from a simple sequence into a highly efficient series of matrix intended to be executed in a CPU or GPU depending on the configuration setting.

To establish a server-based digitized records management the system will utilize the use of OCR tool (see method 3.1) to determine what kind of document is being uploaded at the moment. The system will

then segregate the document's base on their classification. After document classification, the system will upload the document to a certain folder in the server for safe keeping. To obtain this the system will need PHP configurations to upload a file to the server. For this study, the configuration will need large size and max time for uploading.

```

; Maximum allowed size for uploaded files.
upload_max_filesize = 40M

; Must be greater than or equal to upload_max_filesize
post_max_size = 40M

```

Figure 3. Php ini Config

Before transferring the file to the server, the system will get the metadata of the file.

```

$(document).ready(function(){
    $('#input[type="file"]').change(function(e){
        var fileName = e.target.files[0].name;
        alert('The file "' + fileName + '" has been selected.');
```

Figure 4. Getting the File

To make a timeline visualization the system will need a JavaScript and CSS embedded in the codes. The JavaScript needed will be the timeline.js and the timeline.css, a vanilla JavaScript plugin to render a responsive, horizontal/ vertical timeline component from plain HTML. This script is copyright of Mike Collins.

To build a real-time notification the system will need AJAX, jQuery, and Bootstrap. The system will fetch records from the backend of the system. Comparing it to the timelines set, once the case will lapse base on the timeline a notification will be viewed on the user interface.

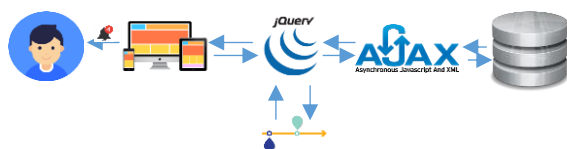


Figure 5. Case Notification

Data Validation and Accuracy To evaluate the system generated output or the text extracted from the document. The

researcher will obtain a copy of the document using word format with the exact set of words as on the extracted text from the document. After extracting the text from the document, the extracted text will then cross-matched with the original text. The number of failed words will be counted and it will be subtracted to the total number of words. The total number of words will serve as the 100% accuracy minus the total words fail it will generate the percent of accuracy. The accuracy test will run through sets of DPI's with ten pages as its parameter. The result will be plotted in excel to generate a graph that will visualize the accuracy difference between the DPI's.

## RESULTS

### Data Extraction from APC Document

Convolutional Neural Network (CNN) is mostly applied to the analysis of visual imagery while Optical Character Recognition (OCR) is the conversion of images into an electronic text. Per the definition of the two, they almost got the same function and used in terms of extracting text from a document like pdf and image file. Base on the researcher's experience, the most convenient to use is the optical character recognition which requires fewer dependencies and system hardware requirements. CNN is implemented with tools like Tensorflow, a free and open source library which is used for machine learning applications. The installation of tensorflow requires python installed on the system, upon training tensorflow requires keras installed on the python. In order to make the training faster, tensorflow requires high-end Graphics Processor Card with CUDA toolkit installed a computing platform model created by Nvidia. While ORC is implemented with the use of tesseract an engine that is considered as one of the most accurate OCR engines available.

The researcher decides to use Tesseract OCR because when it comes to deployment, a system with tesseract engine can be easily

deployed without installing other dependencies. A server-based system with a live running system needs extra care in terms of deploying a new system.

Based on the confusion matrix generated with keras, the accuracy rate for keras falls to 50% with four predictions for CSR while the true value is HSR, predicted 4 CSR with the true value of 4 CSR. While on tesseract OCR the document will be detected based on the extracted text. If the system finds a Case Study Report or Home Study Report it will decide whether the document is a CSR or HSR.

### Data Extraction using Tesseract

The data was extracted from the APC document using tesseract.js, a JavaScript format of Tesseract OCR. The APC document was uploaded to the system and an HTML canvas will hold temporarily the document. The canvas will also serve as a page selection for a multiple page document. The user will select to what page of the document will be extracted. After selecting the page, the user will trigger the extract command to start the extraction process. On the backend of the extraction process, the selected page will be transformed from pdf into an image. The data from the image will be extracted with the use of tesseract, after extracting the text, the results will be split to make new lines using jquery split function. To identify the needed data for data storing will be determined with the use of jquery indexOf function and substring function.

Figure 25 shows the extracted data from the APC document. On the pop-up modal displays a form with corresponding data that was needed to be stored on the database. On the right side is the console where all the extracted data was displayed. The console display is the programmer's determiner if the system runs accordingly.

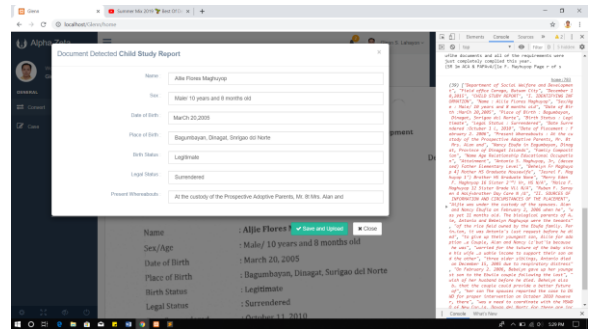


Figure 6 Extracted Data

### Document Classification

Document are classified in order to determine the specific data to be extracted and stored to the database. Document classification follows the data extraction from the APC document. The system is enabled to determine a Child Study Report and Home Study Report. After extracting the data, an argument code was called on a jquery to determine whether the Child Study Report or Home Study Report words were present. With this, the system can determine what type of document was being processed. Figure 26 shows the part of the form where the name of the detected document was shown. The name of the document was automatically detected during the process of data extraction.

Figure 7 shows the code that was utilized by the user in order to classify the document type. In order to classify the document which will hold the whole extracted text from the document. The extracted text will be transformed into the lower case for informality and with the use of the index of function, if the function finds child study report it will concluded that the document is a CSR and if/ however in some cases if the function will find home study report it will now then be concluded that the document is an HSR.

```

function extractText(){
var myImage= document.getElementById('pdf-canvas');
tesseract.recognize(myImage).then(function(result){
var a = result.text;
console.log(result.text);
if (a.toLowerCase().indexOf("child study report") >= 0){
cse(result.text);
}else if(a.toLowerCase().indexOf("home study report") >= 0){
hse(result.text);
}
});
}
function cse(a){
console.log(a.split("\n"));
var lines = a.split("\n");
$.each(lines, function(n, elem) {
if (elem.toLowerCase().indexOf("name :") >= 0){
str = elem.substring(elem.indexOf("-") + 1);
s("aname").empty().val(str);
}
}
}

```

Figure 7. Document Classification

## Server Based APC Digitized Records Management

Digitized APC records refer to the scanned document of APC in the form of pdf or image file. After extracting the data from the APC document, important data will be viewed on the browser for editing and finalization. After finalizing the form, the data will be sent to the database while the document will be uploaded to the server using ajax file upload, the form will be sent to a URL using post request.

## Timeline Visualization of Successful APC

The data that was stored on the database will be selected and retrieve in order base on the id of the case. The retrieved data will be displayed on a table for inventory. With a button for timeline visualization, the case will be displayed in the form of a timeline. Timeline visualization was made using timeline JavaScript. The package contains CSS for timeline styles.

## Case Notification

System notification is in the form of visual notification on the user’s dashboard. This notification will show what is lacking in the case in order for the case manager to respond and comply with the case timelines. The notification was made with the use of font awesome icons with a badge. This badge will display the number of notifications that needs attention. Notification will be automatically form using jquery document ready function with a query regarding on the timeline of the case.

## Data Validation and Accuracy

Figure 8 shows the clip of inaccurate data extracted from the system. On the date of birth, the march word shows the capital C which supposed to be in lower case. The Surigao del Norte words show an n instead of u and on del shows a c instead of e.

Date of Birth :	MarCh 20,2005
Place of Birth :	Bagumbayan, Dinagat, Snrigao del Norte

Figure 8 Data Accuracy

To further test the accuracy of Tesseract OCR, the researcher plots the accuracy rate with ten pages of the document at different dots per inch (DPI). Given DPI are 75, 150 and 300 with a document generated from Microsoft word. To count the number of words the researcher utilizes the Microsoft Word application and compare the extracted text to its original text. Then the failed words were count and subtracted to the total number of words.

Table 1 shows the percent accuracy of the document with 75 DPI. The test document contains 10 pages of the different word found in an HSR. Base on the result of the testing, the average accuracy of 75 DPI is 94.8548778.

Table 1 75 DPI Accuracy Percentage

75 DPI			
Page	Number of Words	Failed Words	Percent Accuracy
1	96	10	89.58333333
2	310	19	93.87096774
3	388	7	98.19587629
4	393	16	95.92875318
5	385	12	96.88311688
6	312	22	92.94871795
7	364	17	95.32967033
8	374	17	95.45454545
9	338	15	95.56213018
10	288	15	94.79166667
<b>Average</b>			<b>94.8548778</b>

The researcher also tested document with 150 DPI resolution. Of all the tested resolution, 150 DPI has the highest average of accuracy which marks up to 99.23619895.

Another test was conducted to a document with 300 DPI. The document has the same content as the mentioned DPI’s before. As shown in table 3, the average accuracy for 300 DPI is 98.91023618.

To have a clean document, the researcher also testes the accuracy to a document that is generated through word application. Table 2 shows the result of accuracy with a clean document. The average accuracy is 98.7416335.

Table 2 Word Processor Generated Result

Page	Word Processor Generated Result			Percent Accuracy
	Number of Words	Failed Words		
1	96	1		98.95833333
2	310	6		98.06451613
3	388	3		99.22680412
4	393	3		99.23664122
5	385	3		99.22077922
6	312	3		99.03846154
7	364	1		99.72527473
8	374	6		98.39572193
9	338	8		97.63313609
10	288	6		97.91666667
<b>Average</b>				<b>98.7416335</b>

During the testing, it is highly noticeable the number of punctuation marks with error in prediction. The result shows that the higher the resolution, the lesser error in punctuation marks.

## CONCLUSION

Based on the results and findings of the study, the text from Alternative Parental Care documents can be extracted using Optical Character Recognition. This text can be filtered to get the significant words needed in the database. APC document can also be classified parallel to the extraction of text from APC document.

The system can also upload and save the document to its server. Large size document will be accepted by the system. The Saved document will be accessible to APC concerned personnel for future use.

Case visualization and notification was also a special feature of the system. The case visualization will be generated based on the data about the cases saved in the database. In connection to the case visualization, a timeline notification was also made to notify the user what is lacking about the case.

Based on the results during testing, the document with 150 DPI resolution scores the most accurate with 99.23619895 %. A Minimal error was noticed during the testing of 150 DPI resolution.

## RECOMMENDATIONS

From the foregoing conclusion, the researcher listed the following recommendation for the improvement of the study:

1. Though Tesseract has already high accuracy rate, it is highly recommended to use other tools or technology that would collaborate Tesseract to obtain much higher accuracy.
2. The researcher also recommends the use of a higher specification computer in order to obtain a faster rendering of image processing results.

## REFERENCES

### References from Journal

Abbott, R., Del Rincon, J. M., Connor, B., & Robertson, N., "Deep object classification in low-resolution LWIR imagery via transfer learning", 2017

Abdul Hamid bin Hamidon, "Real-time mailbox alert system via SMS or email", Aug 4, 2016

Abien Fred M. Agarap, "Deep Learning using Rectified Linear Units (ReLU)", Mar 22, 2018

Alessandro Bissacco, Mark Cummins; Yuval Netzer; Hartmut Neven," PhotoOCR: Reading Text in Uncontrolled Conditions", 2013

Amit Shekhar, "How Does The Machine Learning Library TensorFlow Work?", February 02, 2018

B.U. Kannappanavar; S. T. Rajanikanta, Satish K. Tandur, "Importance Of Digitization Of Library Materials", February 2010

Chirag Patel, Atul Patel, PhD., Dharmendra Patel, "Optical Character Recognition by Open Source OCR



- Tool Tesseract: A Case Study", October 2012
- Christian Bartz Haojin Yang Christoph Meinel, "STN-OCR: A single Neural Network for Text Detection and Text Recognition", July 27, 2017
- Christopher A. Plaisance, "PDF Metadata Extraction with Python", February 5, 2019
- Creating Real-Time Notification System in PHP and AJAX, Jul 4, 2017
- Danqing Liu, "A Practical Guide to ReLU", Nov 30, 2017
- Erick Pierson, "Text Extraction and OCR with Tesseract and Image Magick", December 9, 2015
- George Martsoukos, "Building a Horizontal Timeline With CSS and JavaScript", Apr 27, 2017
- Harald Scheidl, "Build a Handwritten Text Recognition System using TensorFlow", Jul 15, 2018
- Heike Adel and Hinrich Schutze, "Global Normalization of Convolutional Neural Networks for Joint Entity and Relation Classification", Aug 7, 2018
- Henrik Soderlund, "Autonomous email notification and booking management system", 2017.
- Jason Brownlee, "Develop Your First Neural Network in Python With Keras Step-By-Step", May 24, 2016
- Ji Wen, "Recurrent Convolutional Neural Network for Relation Classification"
- Judy J. Johnson, James R. McElroy, Jr, "Computer-Based Records Management System Method", Jul. 28, 1995
- Mike Stark, "Optical Character Recognition with One-Shot Learning, RNN and TensorFlow", 2017
- Lak Lakshmanan, "How to do Text Classification with CNN's, TensorFlow and Word Embedding", Jul 6, 2017
- Leandro S. Guedes, Rafael Garcia, Bruno Pagno, Luciana Nedel, Jo~ao Comba and Carla M.D.S. Freitas, "Interactive Timeline Visualization of Documents", August 2014
- Marc Garcia Torrent, "Install Laravel 5 with Xampp", Feb 3, 2016
- Matt West, "Uploading Files with AJAX"
- Melissa Hedge, "Cloud Storage VS Local Storage – Which is Right for Your Business?", Feb. 3, 2016
- Minh-Thang Luong Michael Kayser Christopher D. Manning Deep Neural Language Models for Machine Translation
- Neeraj Agarwal, "Send Email via Gmail SMTP Server in PHP"
- Nicholas T. Smith, "Deep Learning OCR using TensorFlow and Python", Oct 14, 2017
- Niklas Donges "How to build a Neural Network with Keras", Apr 3, 2018
- Phong H Nguyen, Kai Xu, Rick Walker, and BL William Wong, "TimeSets: Timeline visualization with set relations", October 2015
- PHPMailer complete tutorial with examples: how to send emails with PHP (updated in 2018 with Composer installation), Aug 10, 2018
- Prabhu, "A Simple Tutorial to Classify Images Using TensorFlow—Step by Step Guide", April 27, 2018
- Ray Smith, Ray Smith, Dar-Shyang Lee "Adapting the Tesseract Open Source OCR Engine for Multilingual OCR", July 25, 2009
- Oludare Olaleye, Ayodele Olaniyan, Olalekan Eboda1, Adeleke Awolere," SMS-Based Event Notification System", 2013

Ross Bulat, “Python to PHP Communication – How to Connect to PHP Services using Python”, Oct 24, 2018

S. V. R. Prabhakar, S.V. Manjula Rani, “Benefits and Perspectives of Institutional Repositories In Academic Libraries”, Jan 2018

Sagar Sharma, “Activation Functions in Neural Networks”, Sep 6, 2017

Saima Khan, Dr. Shazia Khan, Mohshina Aftab, “Digitization and its Impact on Economy” June 2015

Sajal Soni, “Upload a File in PHP”, Sep 10, 2018

Sajid, “Send Email Using Gmail SMTP Server from PHP Scrip”, Jan 21, 2019

Sam Deering, “jQuery String Contains Functions”, Mar 29, 2011

Scott LeeTiernan, Edward Cutrell, Mary Czerwinski, and Hunter Hoffman, “Effective Notification Systems Depend on User Trust”

Segit Prasetya Nugroho, “How to make A Web Push Notification in PHP, JQuery, Ajax and MySQL”, Feb 8, 2017

Severino Rebecca, “Chart Combinations: Overlaying Extra Data Onto Timelines”, Dec 22, 2017

Siva Kumar a/l Subramaniam, Siti Huzaimah binti Husin, Yusmarnita binti Yusop,

Sumit Saha “A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way”, Dec 15, 2018

Suncong Zheng, Jiaming Xu, Peng Zhou, Hongyun Bao, Zhenyu Qi, Bo Xu, A neural network framework for relation extraction: Learning entity semantic and relation pattern”, 2016

Swadhin Pradhan, Lili Qiu, Abhinav Parate, and Kyu-Han Kim, “Understanding and Managing Notifications”, 2017

Tao Wang\* David J. Wu\* Adam Coates Andrew Y. Ng, “End-to-End Text Recognition with Convolutional Neural Networks”

TJ Van Toll, “Better Ways of Comparing a JavaScript String to Multiple Values”, Mar 14, 2013

Wesley Yu, “Why the Notification is the Most Important Email You Can Send for Growth”, November 19, 2014

Yash Agarwal, “Create your first Image Recognition Classifier using CNN, Keras and TensorFlow backend”, Jul 8, 2018

### References from Websites

How to train Tensorflow models, Jul 18, 2017. Retrieve from <https://towardsdatascience.com/how-to-traine-tensorflow-models-79426dabd304>

Jamaica B. Lingbanan, “Importance of Digitization and Content Organization”, May 25, 2016. Retrieved from [https://www.slideshare.net/jamaicalingbanan1/importance-of-digitization?from\\_action=save](https://www.slideshare.net/jamaicalingbanan1/importance-of-digitization?from_action=save)

Responsive Horizontal/Vertical Timeline in Vanilla JavaScript – timeline.js, Nov 5, 2018. Retrieve from <https://www.cssscript.com/responsive-horizontal-vertical-timeline/>

Tensors and operations, Retrieve from [https://www.tensorflow.org/js/guide/tensors\\_operations](https://www.tensorflow.org/js/guide/tensors_operations)