

A Generic System for Processing Insurance

Samundeswari S^{a,1}, Jeshoorin G^b, Vasanth M^b,

^{a,b}*Department of CSE, Sri Sairam Engineering College, Chennai, India*

Abstract. Insurance companies are regularly provided with health check reports by the buyers of insurance. Different forms of printed lab reports / health check reports have to be digitized for each value of captured parameters. Optical Character Recognition (OCR), is used to convert the images of handwritten, typed, printed text or any kind of scanned documents into machine-encoded text in order to digitize the values from the report. Conversion to this standard set of digital values will benefit in automating a lot of backend approval process. We collect the reports from the user and read the values from the report and scrutinize the values. Based on the values with the company's standard set, the scrutinization is done and it is then visualized using any visualization tool. The result is presented to the user so that the user can get an idea whether he/she is eligible for insurance claim. The foremost objective of this paper is making the insurance backend approval process a lot easier and a quick response to the buyers.

Keywords. Optical Character Recognition, Insurance, Visualization, digitization.

1. Introduction

Insurance can help the people to protect from loss. It's protection from a variety of risk, primarily made up against the chance of untimely or uncertain loss. Figure 1 shows the number of policies taken by customers taken on various sectors. A person who provides insurance is considered as an insurer, underwriter, insurance carrier or insurance underwriter. Someone or a person who buys insurance is considered as an insured or as a policyholder. The loss or event of loss occur may or might not be of financial manner, but it is converted to financial terms, and often has something with which the insured customer has an interest over by ownership, or possession of the product, or have pre relationship. The insured person is provided with a contract, called the contract, which have a detailed conditions and circumstances under which the insurance company will provide the compensation to the insured person. The amount of cash charged by the insurance company to the customer for the policy from the insurance company is termed as premium.

^bSamundeswari S, Department of CSE, , Sri Sairam Engineering College, Chennai, India
Email: samundeswari.cse@sairam.edu.in

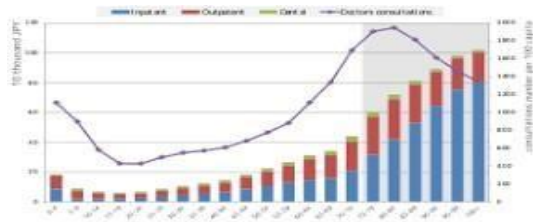


Figure 1. Graph Representing number of policies taken by Customers on Healthcare Sector.

1.1 Scope of project

- i) This method is utilized by any company which has great amount of data and wishes to scrutinize them easily.
- ii) This work also helps the shoppers get their insurance as soon as possible after processing their report.
- iii) This work helps the workers to cut back their work load by scrutinizing the specified reports where human errors will be avoided.

2. Domain Overview

Machine learning involves the detailed study of computer algorithms that plays a major role in automating and improving automation through experience. It's a subdomain of Data Science. Machine learning algorithms (ML) are used to build a mathematical model supported by sample data and training set, remarked as "training data", to create decisions or predictions without writing explicit programs to do. Machine learning algorithms are widely used in a variety type of real time applications, like email filtering, computer vision, where it's very difficult or impossible to develop mathematical algorithms to perform the assigned tasks. Machine learning is closely related to statistics and probability, which is used for prediction using data set. Processing is a field of study, specialized in data analysis through any type of unsupervised learning.

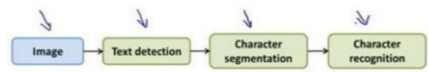


Figure 2. Process Flow of Character separation from images

2.1 Working of OCR

Google’s Tesseract is an open-source text recognition (OCR) Engine. It is either used directly using flask, or using an API to extract texts from images. It supports a wide and all kinds of languages. Google Tesseract doesn't have a built-in user Interface, but there are several available for taking from other sources. Tesseract goes hand in hand with many programming languages and made up of frameworks by using wrappers, Fig 3 shows the working of google Tesseract OCR. It can also be used with the already existing layout analysis from various sources to recognize text from any document.

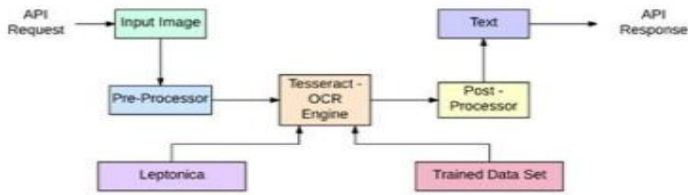


Figure 3. OCR Process Flow

2.2 Constraints

Existing systems are primarily built around one or two core functionalities and fail to integrate all the features that would provide a complete solution towards improving patient comfort and self-reliance at an affordable cost. Some of the early versions of such stature remain costly or inaccessible to the public outside of their places of origin.

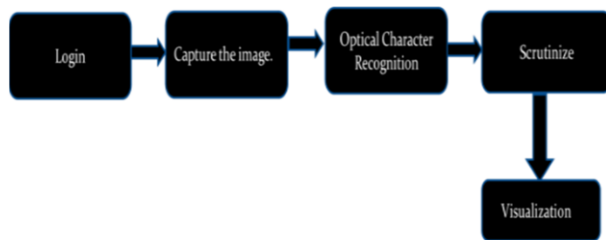


Figure 4. Block view showing the overall system implementation

3. System Implementation

Optical Character Recognition (OCR) which is an electronic or mechanical conversion of images of typed handwritten or printed text into machine encoded text. OCR helps to get the data from the printed lab reports / health check reports to discrete values of each parameter. JSON is a key-value pair output from OCR, which is easy for the Application, to extract the information from the report and send it to the database. The application is developed in such a way where the data received from patient's report is scrutinized and the data is visualized using Google charts.

3.1 Designing The Ui And Uploading The Image To The Firebase

The user Can Using the various Social platform (Google, Facebook....) provided on the login page. The packages package:./dart and package:./dart are imported for the above task. Once the sign in takes place two tokens are returned as a result ("Access token" and "ID Token"). Using the Access Token, the information of the user is taken from the user's profile on the signed in social account. Then the details are stored for future purposes. The data of the user stored. If the user chooses to upload his profile pic it is



Figure 5. User Interfaces showing login, profile and image upload screen

stored in the application. Claim pages is where the user gets the status of his/her insurance processing status can also view the reports uploaded by the user, since the usage of application in reports page Figure 6 shows the User Interfaces of login, profile and image upload screens.

3.2 *Getting Values From The Report*

The link generated is then passed on to a private OCR by using API. Newer Version of Tesseract supports a new neural net-based engine which is all about line recognition and used in paper [2][3][5]. Tesseract also has Unicode support, and can recognize more than n number of languages. Tesseract provides supports for various output formats such as: Portable Document Format (PDF), plain text, invisible text PDF, Tab Separated Values (TSV). The support for ALTO (XML) output is just experimental as provided by Optical Character recognition (OCR) tool present in python. what it means is that it is able to recognize the text embedded or present in images. Python tesseract is a type of wrapper using Google's Tesseract OCR Engine. It is always a pretty useful stand-alone invocation script tool for invocation of tesseract, since it is able to read all images supported by the Pillow and imaging.

3.3 *Scrutinizing, Visualising The Data And Backend Processing*

The incoming data is accessed using Flask, and have to use the request object. The request object gets holds of all the incoming data from the request, which also includes the mime type, referrer, raw data, HTTP method, and even headers. For easily gaining access to the request object provided to the Flask, we can simply import it from Flask library. The Result is then stored in a variable called reply from which u can get the data and scrutinize the data based on needs. The Sync fusion Flutter Charts package is one of the data-visualization libraries written in Dart for creating a beautiful and high-performance chart which is also used in the paper [6][9]. Figure 6 shows a sample screen of the results, which is obtained from the medical report. Based on the data, we initialize the appropriate axis type and series type.



Figure 6. Results obtained from the medical report

4. Conclusion

This paper outlined the practical implementation of the Digitization of insurance approval by which a lot of time can be saved. This method can be used by any company which has large amount of data and needs to scrutinize them based on the conditions. This project also helps the customers get their insurance as soon as possible after scrutinizing their report. This project helps the employees to reduce their work load by scrutinizing the required reports where human errors can be avoided.

References

- [1] A. C. Downton, A. Amiri, L. Du and S. M. Lucas, "A configurable toolkit approach to handwritten forms recognition," IEE Colloquium on Document Image Processing and Multimedia Environments, London, UK
- [2] Adapting the Tesseract Open Source OCR Engine for Multilingual OCR Ray Smith, Daria Antonova, Dar Shyang Lee Google Inc., USA.
- [3] Unnikrishnan, R., Smith, R., Combined Script and Page Orientation Estimation using the Tesseract OCR engine Submitted to International Workshop of Multilingual OCR, 25th July 2009, Barcelona, Spain.
- [4] Creation of custom recognition profiles for historical documents Adam, and Tomasz Parkoła Poznań and Networking Center.
- [5] An Overview of the Tesseract OCR Engine, Ray Smith, Google Inc.
- [6] An Interactive Document Image Description for OCR of Handwritten Forms David Monger, Graham Leedham and Andy Downton University of Essex, United Kingdom.
- [7] G. Leedham and D. Monger, "Evaluation of an interactive tool for handwritten form description," Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, Quebec, Canada.
- [8] Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments Katy ,1, and Michael a School of Informatics, Computing, and Engineering, Indiana University,
- [9] 1. Lee PS, West JD, Howe B (2018) Vizometrics: Analyzing Visual Information in the Scientific Literature. IEEE Transactions on Big Data 4(1):117–129.
- [10] Borkin MA, et al. (2016) Beyond Memorability: Visualization Recognition and Recall. IEEE Trans Vis Comput Graph 22(1):519–528.
- [11] Gingerich O, van Helden A (2003) From Occhiale to printed page: the making of Galileo's Sidereus nuncius. Journal for the History of Astronomy 34:251–267.
- [12] P. V. Klaine, M. A. Imran, O. Onireti and R. D. Souza, A Survey of Machine Learning Techniques Applied to Self Organizing Cellular Networks, in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, 2018.
- [13] B. Akshaya, M. Rajendiran. Securing Documents Using Digital Certificate-A Review. International Journal of Research in Advent Technology (IJRAT), 2018
- [14] M. Dhivya, K. Valarmathi. IoT Based Smart Electric Meter. Lecture Notes on Data Engineering and Communication Technologies, volume 35, pp 1260-1269, 2020.
- [15] R. Vidhya, K. Valarmathi. Automatic Monitoring of Hydroponics System Using IoT. Lecture Notes on Data Engineering and Communication Technologies, volume 35, pp 641-648, 2020
- [16] Rajamani.V, Murugavalli.S. A High Speed Parallel Fuzzy C-means algorithm for Tumor Segmentation. ICGST International Journal on BIME 6 (1), 2006