**BABEȘ-BOLYAI UNIVERSITY CLUJ-NAPOCA**

**FACULTY OF MATHEMATICS AND COMPUTER**

**SCIENCE**

**SPECIALIZATION COMPUTER SCIENCE**

**DIPLOMA THESIS**

Car registration by document classification using

computer vision

**Supervisor**

Asist. dr. Petrescu Manuela

**Author**

Pop David Alexandru

**2024**

**UNIVERSITATEA BABEȘ-BOLYAI CLUJ-NAPOCA**

**FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ**

**SPECIALIZAREA INFORMATICĂ**

**LUCRARE DE LICENȚĂ**

Inmatriculare autovehicul prin clasificare de documente folosind viziune computerizata

**Supervisor**

Asist. dr. Petrescu Manuela

**Author**

Pop David Alexandru

**2024**

Table of Contents

[1. Introduction 5](#_Toc159444708)

[2. Existing methods for document classification 6](#_Toc159444709)

[2.1 Deep learning 6](#_Toc159444710)

# 1. Introduction

In an era where the volume of digital documents is expanding exponentially, the need for effective organization and classification methods is paramount. Document classification, the task of automatically assigning predefined categories to text documents, plays a pivotal role in numerous real-world applications such as registering a car. The main problem in this thesis is classifying documents using computer vision. Generally, there are two solutions: manual classification and automatic classification. The manual approach implies that the user specifies the type of document. The automatic approach makes use of the technology Optical Character Recognition, known as OCR. This approach does not come without problems. Text extraction is made using OCR engines. Such an engine is built to recognize characters in various fonts and sizes. Problems which might arise when using an OCR engine could be characters are not textual, hence using an OCR engine on an image is not enough to get the text. This system needs to consider different transformations to have better accuracy for character recognition. Such kind of transformations are: different qualities, blurring, warp perspective, rotating. The proposed solution warps the image and apply the above-mentioned transformations to get the text match and find the document from given templates. Another problem that arises is data privacy. The system uses all the privacy configuration needed from password hashing, session management up to image persistence on cloud services. All these features can be used using an android app interface written in Dart using the Flutter framework. The focus of this thesis is the system behind it which is exposed as an API written in object-oriented language, namely C#. Document classification microservice is written in Python using the OpenCV library. The technique use to classify documents will be presented in the third chapter. The explanation and trade-offs with advantages and disadvantages against other technologies and architecture decisions are presented in the fourth chapter.

# Existing methods for document classification

In this section, we will explore two approaches of how we can classify documents. Firstly, we will review a machine learning method based on deep learning, then a computer vision method based on OCR and pattern matching using regular expressions.

## **Machine Learning**

Logistic regression is one of the known and used supervised learning algorithms that are used to indicate the probability of a category is logistic regression. Given a document d represented by a set of features , the probability that P( belongs to a class is calculated as follows:

P = σ()

A text classifier is implemented in several steps. In the first step, labeled documents are represented as vectors of a certain length. The documents will be then filtered using data sanitization such as removing unwanted characters, punctuation and symbols. Furthermore, the documents will be divided into two sets, the training set and the test set. The training set will used to fed into classifiers to train them, while the test set will be used to evaluate and predict the results [1].

## **Computer Vision and pattern matching**

Moden OCRs make use of neural networks (e.g. PaddleOCR, Tesseract), these being trained, for example, to recognize entire lines instead of single characters. The proposed solution for this thesis is to determine the set of all contours of the image. From the contours set, we extract the biggest contour, determined by its area, and warp perspective of the contour with all our templates so that is perfectly aligned with the template. Following this, a serios of different image transformations are applied to the biggest contour such as blurring and grayscale. Given that certain documents, such as identity cards, may contain sensitive facial information, appropriate measures are taken to obscure these facial features utilizing the Haar-cascade algorithm [2]. To see how much looks like the current template we will use a quality metric, RASE (Relative Average Spectral Score) and save the best match, having the smallest RASE [3]. The motivation behind using this quality metric is, first of all, speed of computation and second of all, reliability, computing a good approximation if two images look the same. After all epochs, we will use an OCR engine to get the text from the best match with the current template and parse the text with a list of predefined regular expressions which have associated a confidence level. Finally, if the confidence level is greater than a threshold then we know the type of document.