

**DETECTION OF FRAUD INVOICES USING IMAGE
CLASSIFICATION
CASE STUDY ON HEALTHCARE INVOICES**

U.D. Nipuni Awarjana Sathsarani Perera

A dissertation submitted in partial fulfillment of the requirement for
Master of Science degree in Business Analytics

Department of Business

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University, UK

2022

Abstract

The insurance organizations, need to manage invoices, and transactions regarding healthcare insurance, among others. Customers those who are insurance holders need to hand over the invoices to the agents of the insurance company, after that agents need to verify the invoices based on the given details. Most insurance companies are verifying invoices using a manual process. This is a challenge for the organization because needs to assign a specifically trained person to invoice verification

This study is mainly focused to introduce machine learning algorithms that can be adapted to an automated invoice process. Before including machine learning models most important phase is the image processing method. This research has four phases. The first phase is collecting invoices data, the second phase is data cleaning and categorizing, the third phase is Image classification model and image processing implementation, and the fourth phase is implementing machine learning models. Invoices image data has been categorized as doctor appointment invoices, pharmacy invoices, and hospital bills.

The image processing used multiple techniques to process the images before integrating with the machine learning models. Such as gray scaling, resizing, and data normalization. After pre-processing the images were used cross-validation for the best results.

During the cross-validation in the first iteration, the first group of the five groups was assigned as the Test data, and the rest of the four groups were assigned as train data. Then train datasets have been grouped and flattened. As discussed in the above image pre-processing method has been used during this step for each image in both valid and invalid invoice image data.

As the final stage implemented the machine learning models and used Random Forest, Decision Tree, KNN, Naïve Bayes, Support Vector Machine, and Artificial Neural Network algorithms to analyze the accuracy and choose the best model for the invoice fraud detecting process.

In the analysis phase of the study has been shown Artificial Neural Network has good performance and all the other machine learning models have more 0.5 accuracy rate. Hence this study has indicated that machine learning algorithms can be adapted for the fraud invoice detection process successfully. Insurance companies can adopt a machine learning model with a high accuracy rate this for their health invoice claiming process.