ORDER IN DOCUMENT CHAOS: LOGISTICS DOCUMENTS CLASSIFICATION

VYTAUTAS MAGNUS UNIVERSITY
Faculty of Informatics

AUTHORS:

Danylo Abramov
danylo.abramov@vdu.lt

Eimantas Zaranka
eimantas.zaranka@vdu.lt

Monika Zdanavičiūtė
monika.zdanaviciute@vdu.lt

Nerijus Šakinis
nerijus.sakinis@vdu.lt

Tomas Krilavičius
tomas.krilavicius@vdu.lt

Co-funded by the European Union

SustAIn Liv Work

terra IT

CARD
CENTRE FOR APPLIED RESEARCH AND DEVELOPMENT

MOTIVATION

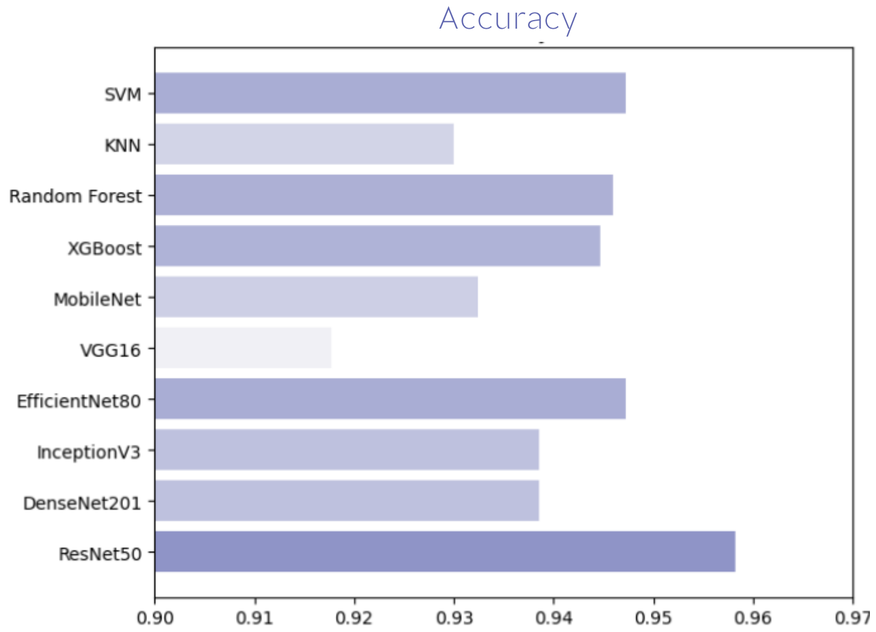
- The global logistics industry, growing by approximately $0.5 trillion yearly, faces an increasing number of documents that need efficient processing.
- Manual handling of logistics documents is time-consuming and can lead to mistakes, underscoring the need for automated document classification.

GOAL

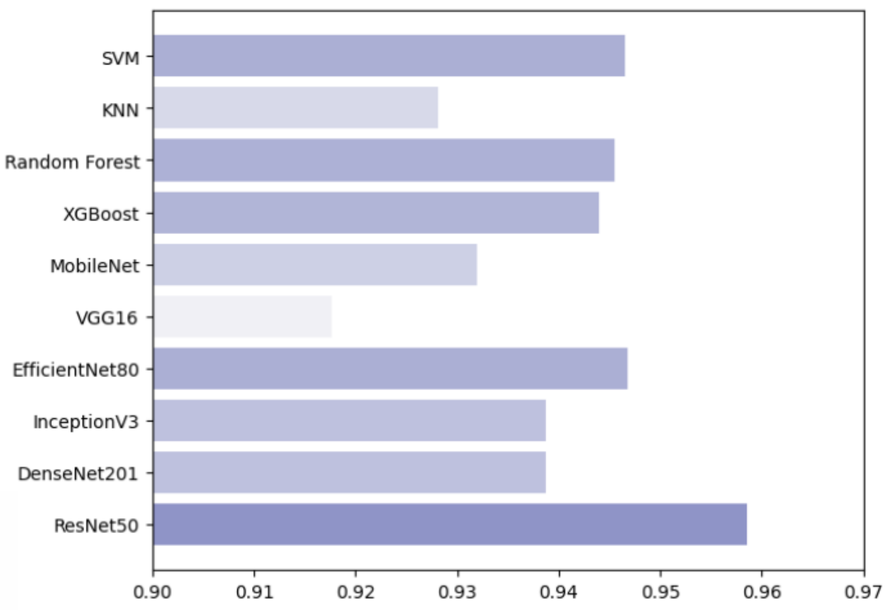
- Develop an efficient classification system for logistics documents: CMRs, Invoices, Receipts, and Others, using deep learning and machine learning techniques.
- Implement the best model to help logistics companies manage documents more efficiently with minimal manual work.

EVALUATIONS

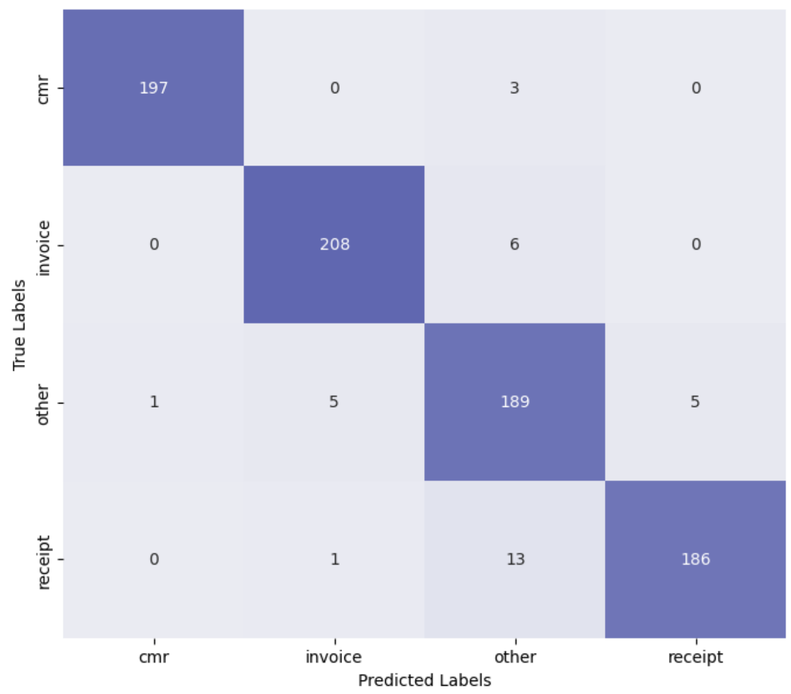
Accuracy



F1 score

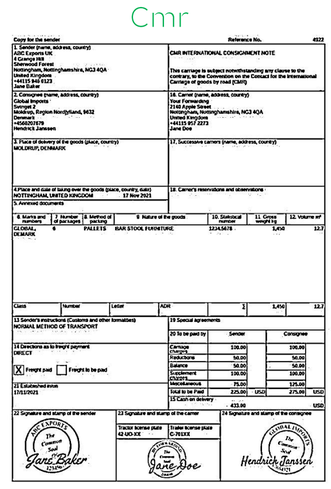


Confusion matrix for ResNet50

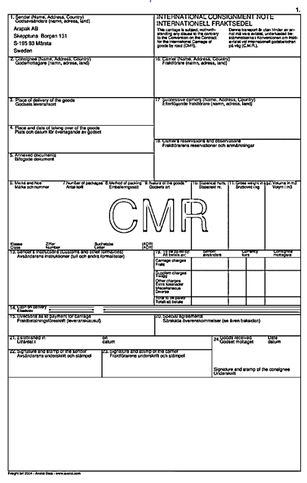


PREDICTIONS

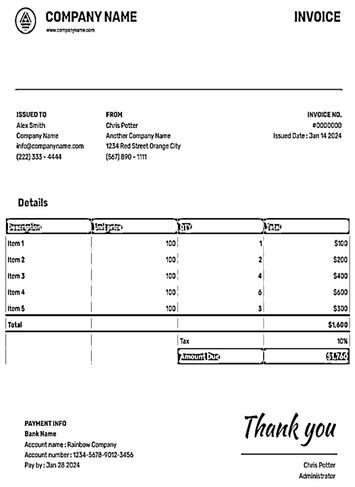
Cmr



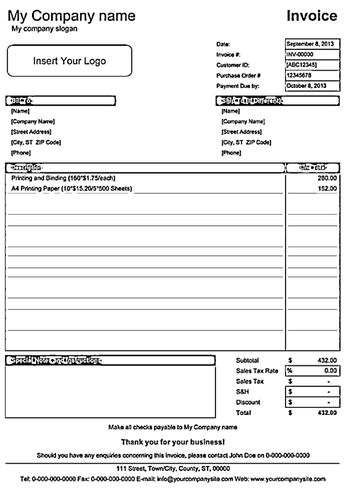
Cmr / Other



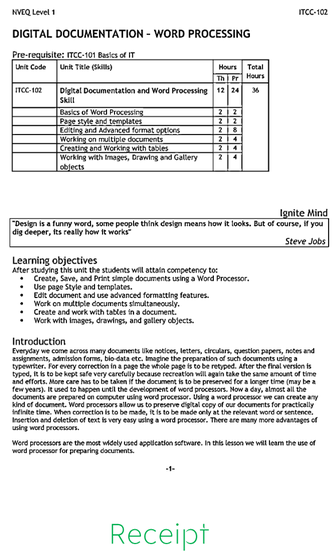
Invoice



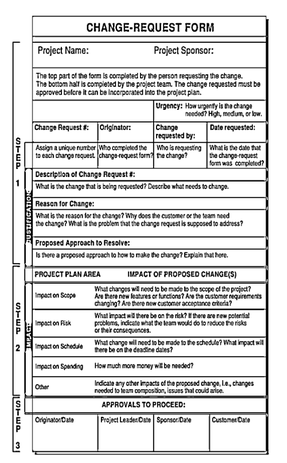
Invoice / Other



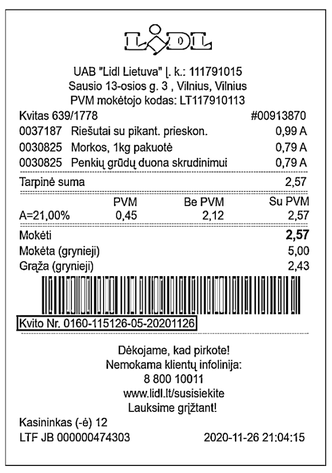
Other



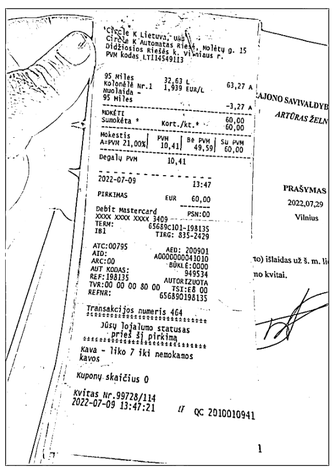
Other



Receipt



Receipt / Invoice



CONCLUSIONS

- The ResNet50-based neural network model achieved the highest accuracy among all tested models, with an F1 score of 0.9585, making it the most effective for logistics document classification.
- Support Vector Classifier was the top-performing traditional machine learning model, achieving an F1 score of 0.9466 using ResNet50 features.
- Future efforts will focus on expanding the dataset and further improving model performance.