Enhancing Multiclass Document Classification with LayoutLM: A Comparative Study of V1 and 2

Kounen Fathima*, Ali Athar*, Hee-Cheol Kim*

Institute of Digital Anti-Aging Healthcare, Inje University, South Korea
kounenfathima00@gmail.com, ali.athar1401@gmail.com, heeki@inje.ac.kr

***Abstract*— In the age of information explosion, efficient and accurate information retrieval has become a pivotal task across numerous domains, from finance to healthcare and beyond. The LayoutLM model has enhanced the capabilities of existing NLP techniques by enabling them to extract information from complexly structured documents automatically. In this study, we employ a subset of the RVL-CDIP (Ryerson Vision Lab Complex Document Information Processing) dataset, consisting of 400,000 pictures of data organized into 16 groups. Out of the 16 classes, this subset includes all the classes but there was a limit set to 200 images per class which makes the total amount of images as 3200. Accordingly, 1920, 640, and 640 images make up the training, validation, and testing sets. This dataset is a rich collection of documents with diverse structures and content to demonstrate the effectiveness of our proposed method. LayoutLM, with its unique capability to analyze and understand document structures, has been a pivotal component of our methodology. We utilized LayoutLMv2 and version 1 for the purpose of classifying the documents into their respective categories and comparing the results accordingly. Accordingly, the two versions' accuracies are 80.94 and 68.75 percents.**

***Keywords*— Natural Language Processing (NLP), LayoutLM Model, Document AI, Document Classification, Transformers**

Kounen Fathima received her BE in IT Engineering from the Osmania University, India. Currently, she is pursuing her master's degree in the Institute of Digital Anti-Aging Healthcare from Inje University. Her research interest aligns with Artificial Intelligence, Machine learning and Natural Language Processing

Ali Athar received his BSSE degree Software Engineering from Government College University Faisalabad (GCUF), Pakistan. He received his MS degree from NUST, Pakistan. He is pursuing his Ph.D. degree from the Institute of Digital Anti-aging and healthcare at Inje University. His research areas include Text Mining, Machine learning, and Deep learning

Hee-Cheol Kim BSc at the Department of Mathematics, MSc at the Department of Computer Science at SoGang University in Korea, and Ph.D. in Numerical Analysis and Computing Science, at Stockholm University in Sweden. He is a professor and Head of the Department of the Institute. Digital Anti-aging Healthcare, Inje University, S: Korea. His research interests include Machine learning, Text mining, and Bio Informatics.