**Logo, company name

Description automatically generated Department of IT & Computer Science**

**Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur, Pakistan**

**COMP-340 Artificial Intelligence**

**Project Proposal**

**Group Members:**

* Jamal Ahmed Khan

(B20F0358CS022)

* Ahmed Raza

(B20F0436CS031)

**Date:**

14th April 2023

**Submitted To:**

Dr. Khalil Khan

# **Abstract**

Handwritten character recognition is a challenging task in natural language processing with numerous applications in various fields such as education, health, banking, and administration. Despite the large amount of written material in Hindko language, there is still no mature OCR system available for it. Therefore, we propose a long-term research strategy to develop a complete OCR system for Hindko. In this paper, we present our initial step towards this goal, which is the development of the first Hindko words dataset and its recognition using deep learning algorithms. Our research objectives are to compile the handwriting samples of Hindko language students and use deep learning methods for feature extraction and classification of these words. This work has the potential to contribute to the scientific community and society by digitizing and analyzing large amounts of handwritten documents, improving accessibility for people with disabilities, and potentially uncovering new insights and patterns.

# **Introduction**

Handwritten text recognition is an active area of research with numerous applications in professional and industrial fields. The first OCR system was developed for Latin text recognition in the late 1950s, and nowadays these systems can recognize a variety of scripts and languages. However, there has been a lack of research effort in Hindko handwritten text recognition, and no dataset exists for this language. Therefore, in this paper, we propose to build a dataset for Hindko language and use deep learning for training and predicting outcomes.

# **Related Work**

Research into OCR systems for various languages, including English, French, Chinese, and Arabic, has been ongoing for many years. However, there is a lack of research into Hindko handwritten text recognition, and no dataset exists for this language. Our proposed work will address this gap in the literature.

# **Proposed Methodology**

Our research objectives are to compile the handwriting samples of Hindko language students and use deep learning methods for feature extraction and classification of these words. To achieve our first objective, we will take handwriting samples from different students at different universities or institutes and compile them into a complete dataset. For the second objective, we will use deep learning-based methods for feature extraction and classification of these words.

# **Conclusion and Future Work**

In conclusion, we propose a long-term research strategy to develop a complete OCR system for Hindko language. As an initial step towards this goal, we will develop the first Hindko words dataset and use deep learning algorithms for its recognition. The potential impact of this work is significant, as it can contribute to digitizing and analyzing large amounts of handwritten documents, improving accessibility for people with disabilities, and potentially uncovering new insights and patterns. In future work, we plan to expand our dataset to include more words and develop a complete OCR system for Hindko.