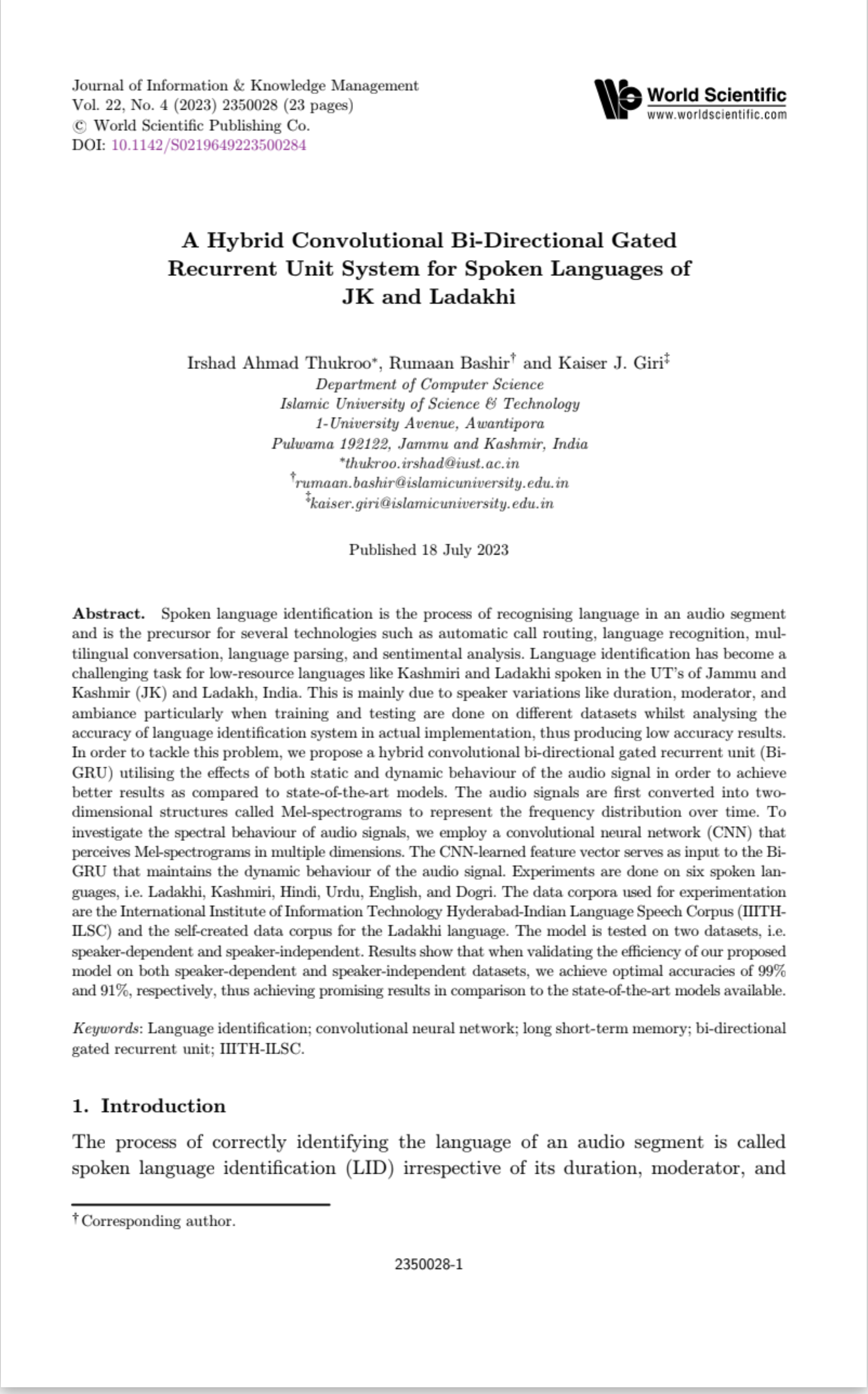

Journal of Information & Knowledge Management
Vol. 22, No. 4 (2023) 2350028 (23 pages)
© World Scientific Publishing Co.
DOI: 10.1142/S0219649223500284

A Hybrid Convolutional Bi-Directional Gated Recurrent Unit System for Spoken Languages of JK and Ladakhi

Irshad Ahmad Thukroo*, Rumaan Bashir† and Kaiser J. Giri‡
Department of Computer Science
Islamic University of Science & Technology
1-University Avenue, Awantipora
Pulwama 192122, Jammu and Kashmir, India
**thukroo.irshad@iust.ac.in*
†*rumaan.bashir@islamicuniversity.edu.in*
‡*kaiser.giri@islamicuniversity.edu.in*

Published 18 July 2023

Abstract. Spoken language identification is the process of recognising language in an audio segment and is the precursor for several technologies such as automatic call routing, language recognition, multilingual conversation, language parsing, and sentimental analysis. Language identification has become a challenging task for low-resource languages like Kashmiri and Ladakhi spoken in the UT's of Jammu and Kashmir (JK) and Ladakh, India. This is mainly due to speaker variations like duration, moderator, and ambiance particularly when training and testing are done on different datasets whilst analysing the accuracy of language identification system in actual implementation, thus producing low accuracy results. In order to tackle this problem, we propose a hybrid convolutional bi-directional gated recurrent unit (Bi-GRU) utilising the effects of both static and dynamic behaviour of the audio signal in order to achieve better results as compared to state-of-the-art models. The audio signals are first converted into two-dimensional structures called Mel-spectrograms to represent the frequency distribution over time. To investigate the spectral behaviour of audio signals, we employ a convolutional neural network (CNN) that perceives Mel-spectrograms in multiple dimensions. The CNN-learned feature vector serves as input to the Bi-GRU that maintains the dynamic behaviour of the audio signal. Experiments are done on six spoken languages, i.e. Ladakhi, Kashmiri, Hindi, Urdu, English, and Dogri. The data corpora used for experimentation are the International Institute of Information Technology Hyderabad-Indian Language Speech Corpus (IIITH-ILSC) and the self-created data corpus for the Ladakhi language. The model is tested on two datasets, i.e. speaker-dependent and speaker-independent. Results show that when validating the efficiency of our proposed model on both speaker-dependent and speaker-independent datasets, we achieve optimal accuracies of 99% and 91%, respectively, thus achieving promising results in comparison to the state-of-the-art models available.

Keywords: Language identification; convolutional neural network; long short-term memory; bi-directional gated recurrent unit; IIITH-ILSC.

1. Introduction

The process of correctly identifying the language of an audio segment is called spoken language identification (LID) irrespective of its duration, moderator, and

† Corresponding author.