A deep learning approach for ncRNA sequences classification

Antonino Fiannaca, Massimo La Rosa, Laura La Paglia, Riccardo Rizzo, and Alfonso Urso

{antonino.fiannaca,massimo.larosa,laura.lapaglia,riccardo.rizzo,alfonso.urso}@icar.cnr.it

National Research Council of Italy
Istituto di Calcolo e Reti ad Alte Prestazioni, Palermo

Motivation

ncRNAs are small non-coding sequences and recent studies have demonstrated they are involved in different biological processes, including diseases. There are several kinds of ncRNAs, which differ each other on the basis of their length, structure (folding) and function.

In this context, it is fundamental to investigate a computational method that can provide identification and classification of different kinds of ncRNA molecules.

In this work we propose a pipeline for classification of ncRNA sequences based on structural features extracted from RNA secondary structure and a deep learning architecture implementing a CNN (convolutional neural network).

Methods
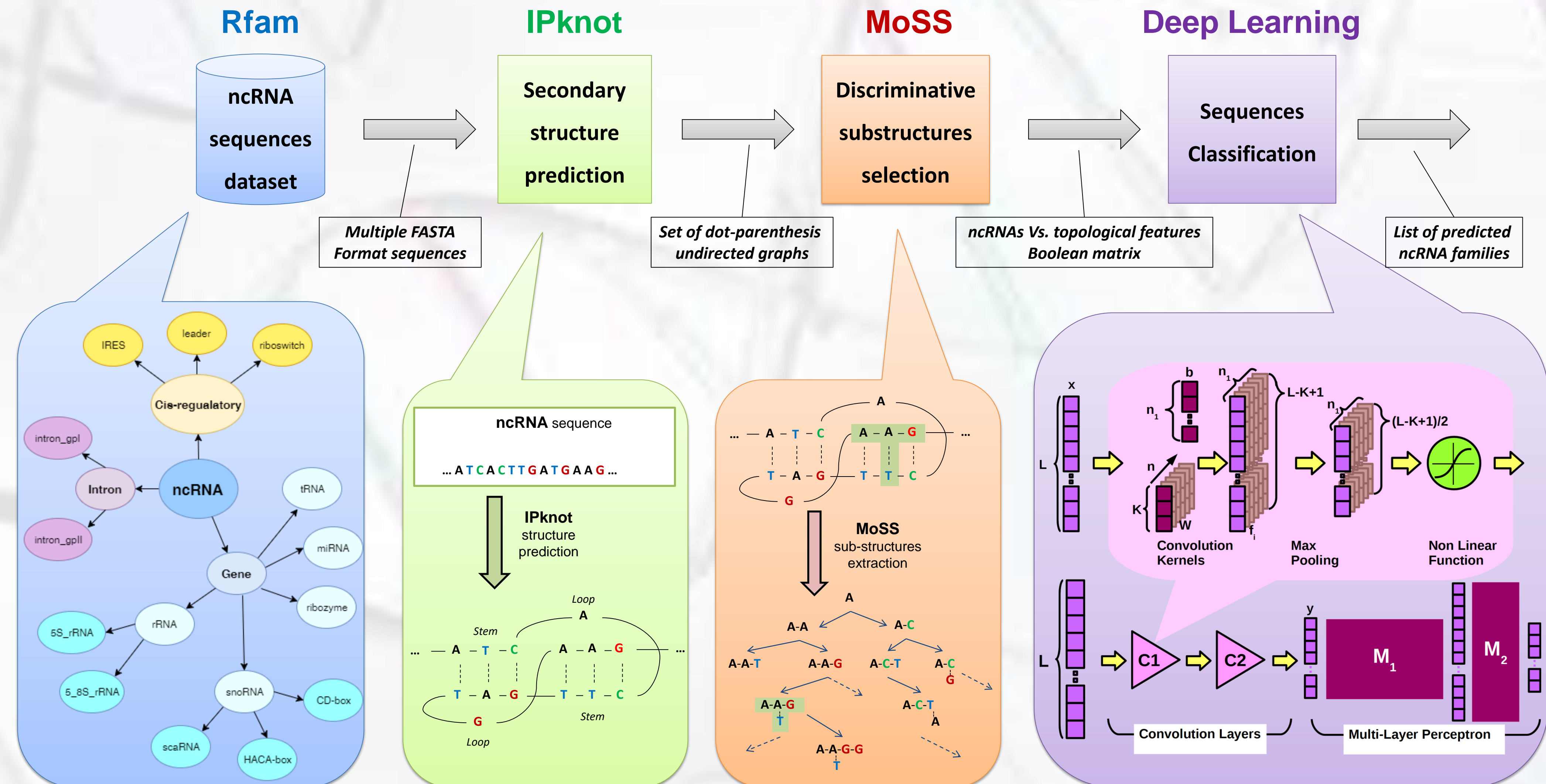
Experimental Results

Algorithm	Accuracy		Sensitivity		Specificity		Precision		F-score	
	%	σ	%	σ	%	σ	%	σ	%	σ
CNN	74.69	0.013	74.13	0.010	97.89	0.001	74.09	0.010	74.11	0.010
RF	56.60	1.544	56.50	1.539	96.41	0.001	55.69	0.019	56.10	0.037
NB	47.31	1.837	48.98	1.844	95.61	0.001	46.80	0.019	47.87	0.037
kNN	54.70	1.609	59.17	1.783	96.28	0.001	54.26	0.018	56.61	0.035
SVM	67.36	1.991	67.47	1.855	97.11	0.001	73.92	0.673	67.76	0.019
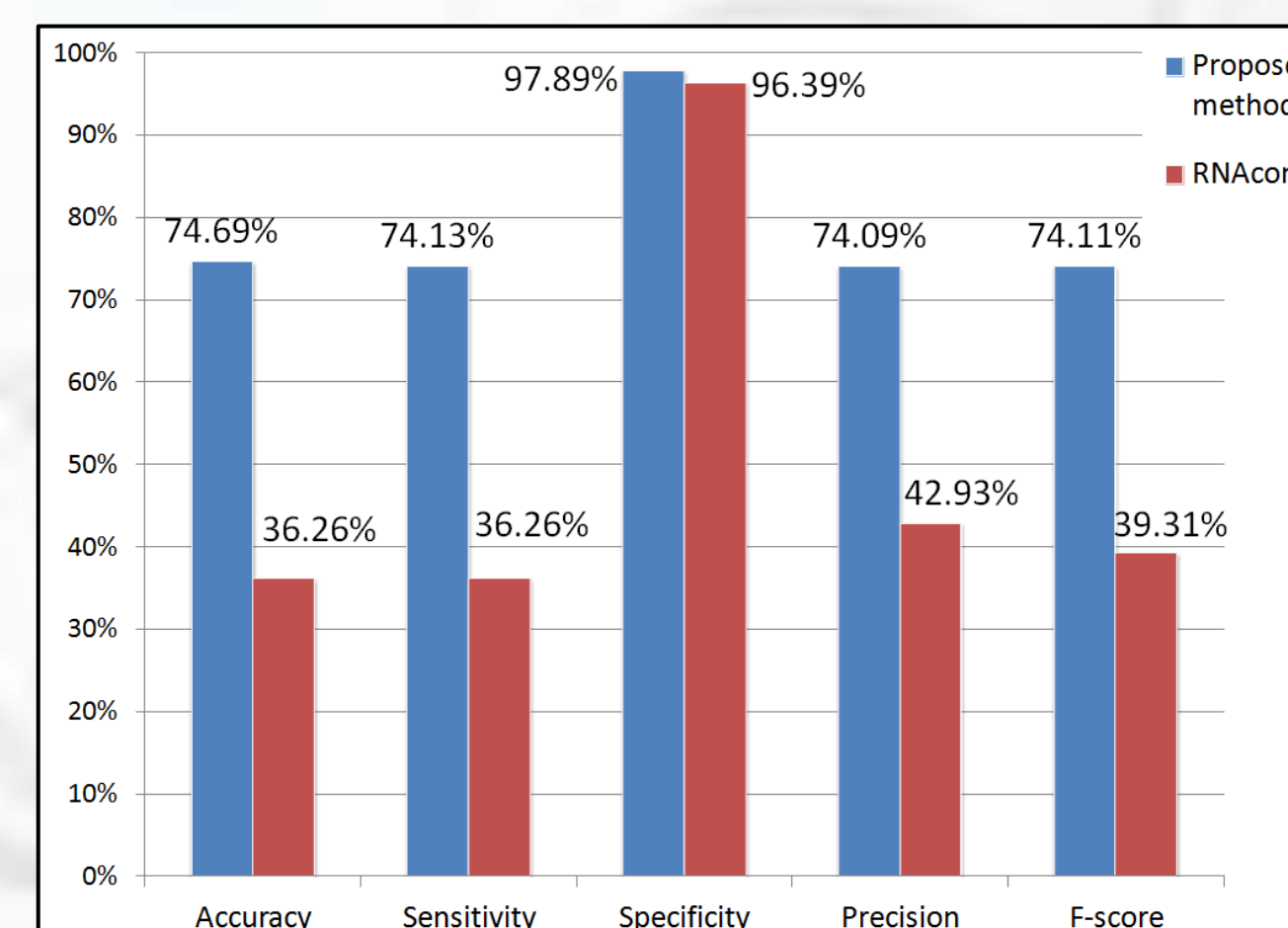
➢ 6300 sequences from 13 ncRNA classes;

➢ Statistical measures over 5 classifiers;

➢ Comparison with RNACon, the state-of-the-art technique for ncRNA classification.