

Beyond the Horizon: Using Mixture of Experts for Domain Agnostic Fake News Detection

Carmela Comito¹, Massimo Guarascio¹, **Angelica Liguori**¹, Giuseppe Manco¹, Francesco Sergio Pisani¹

¹ Institute for High Performance Computing and Networking, National Research Council, (ICAR-CNR), Italy



Discovery Science 2024
October 14-16 2024 | Pisa, Italy



Outline

- I. Problem Overview
- II. Proposed Approach
- III. Experimental Assessment
- IV. Conclusion and Future Directions

Social Media & Fake News

- Unverified information shared through social networks could contain malicious/false content
- Such fake news can span across different domains
 - *Uncovering fake news on social media is decisive and challenging*



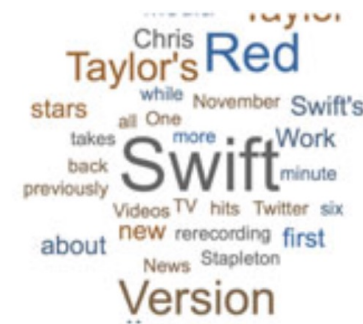
Fake News Detection Challenges

- Fake news generation rate dramatically grows when disruptive events happen
 - Lack of labeled data
- Fake news detection models, typically, are trained on specific domains
 - They exhibit poor performance in the detection of fake news in new domains

Domain-specific word usage



Source Domain D_s



Target Domain D_t

SoTA Limitations

- Deep learning-based cross-domain fake news detection SoTA
 - Requires knowledge of the target domain
 - Not able to catch new emerging events
 - Requires a substantial number of labeled examples from the target domain
- Fake news detection on **new** (unseen) and/or emergent **domains**
 - Leverage knowledge from known domains to effectively identify fake news in unfamiliar contexts

Methodology Overview

- **Goal.** Devising a cross-domain fake news detection model based on deep learning architectures for revealing false/unverified information on unknown domains
- **Idea.** Combining language models with ensemble strategies
- **Solution.** Defining **MoDA – FND** (Mixture of Domain Agnostic Fake News Detectors), a hierarchical deep ensemble approach for learning cross-domain fake news detectors from text data

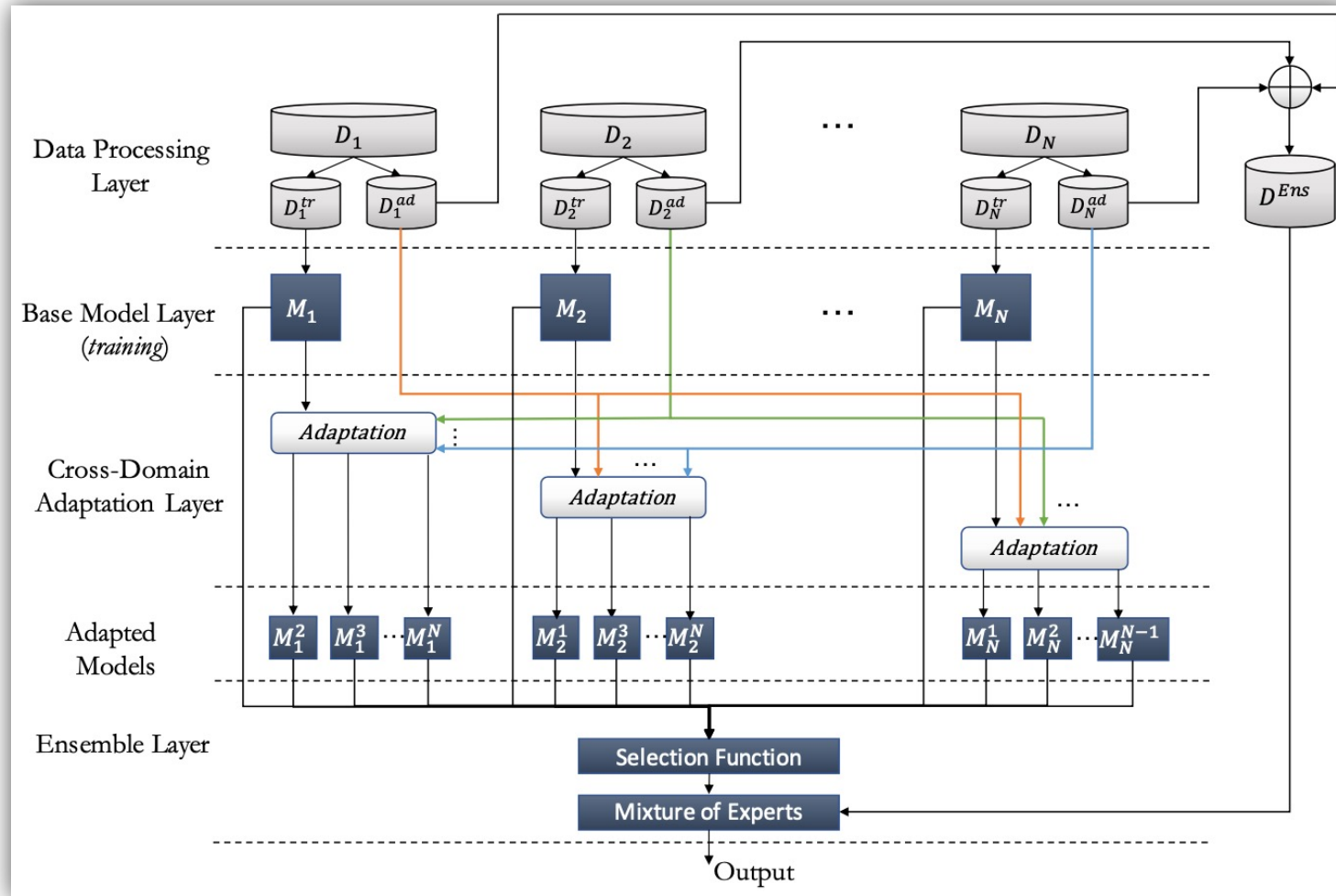
Problem Definition

Statement

Learn a function f from a set of known different application domains (source data) $\{D_1, D_2, \dots, D_N\}$ that can predict the real label of any news from an unknown domain (target data) D_T

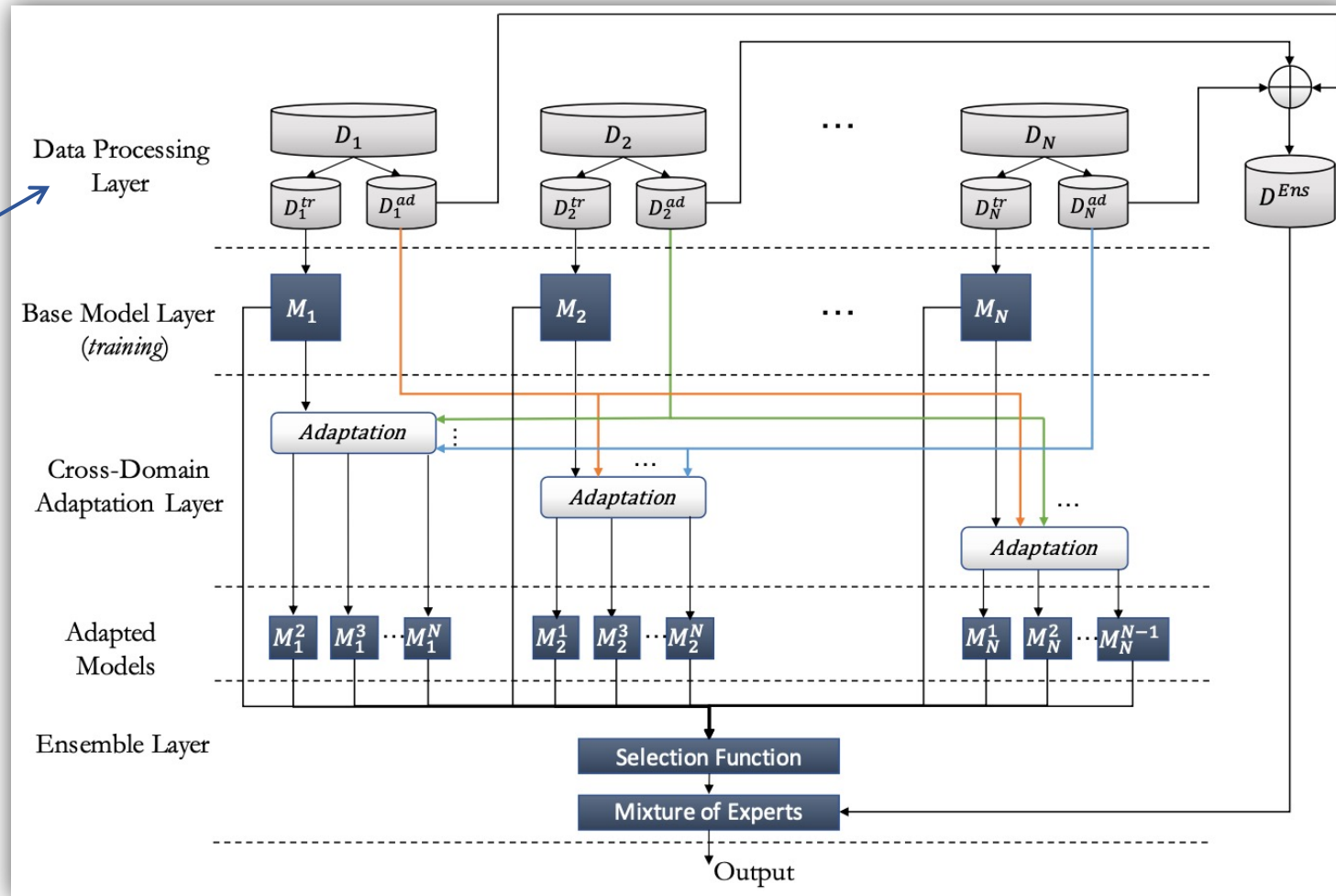
In our setting, f takes the form of an ensemble of DNNs

Proposed Solution: MoDA - FND



Proposed Solution: MoDA - FND

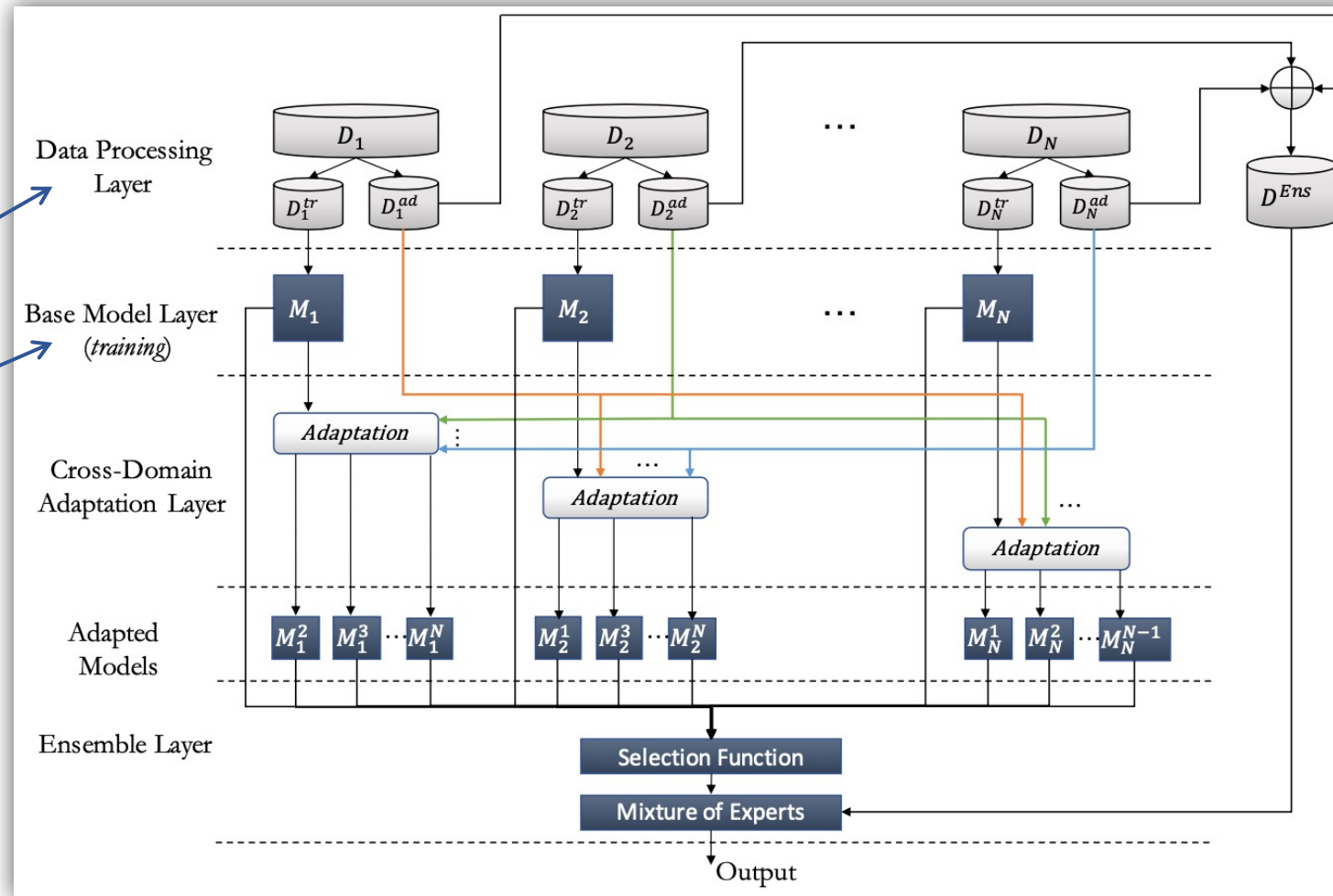
1 Split datasets



Proposed Solution: MoDA - FND

1 Split datasets

2 Train a base specialized fake news detector

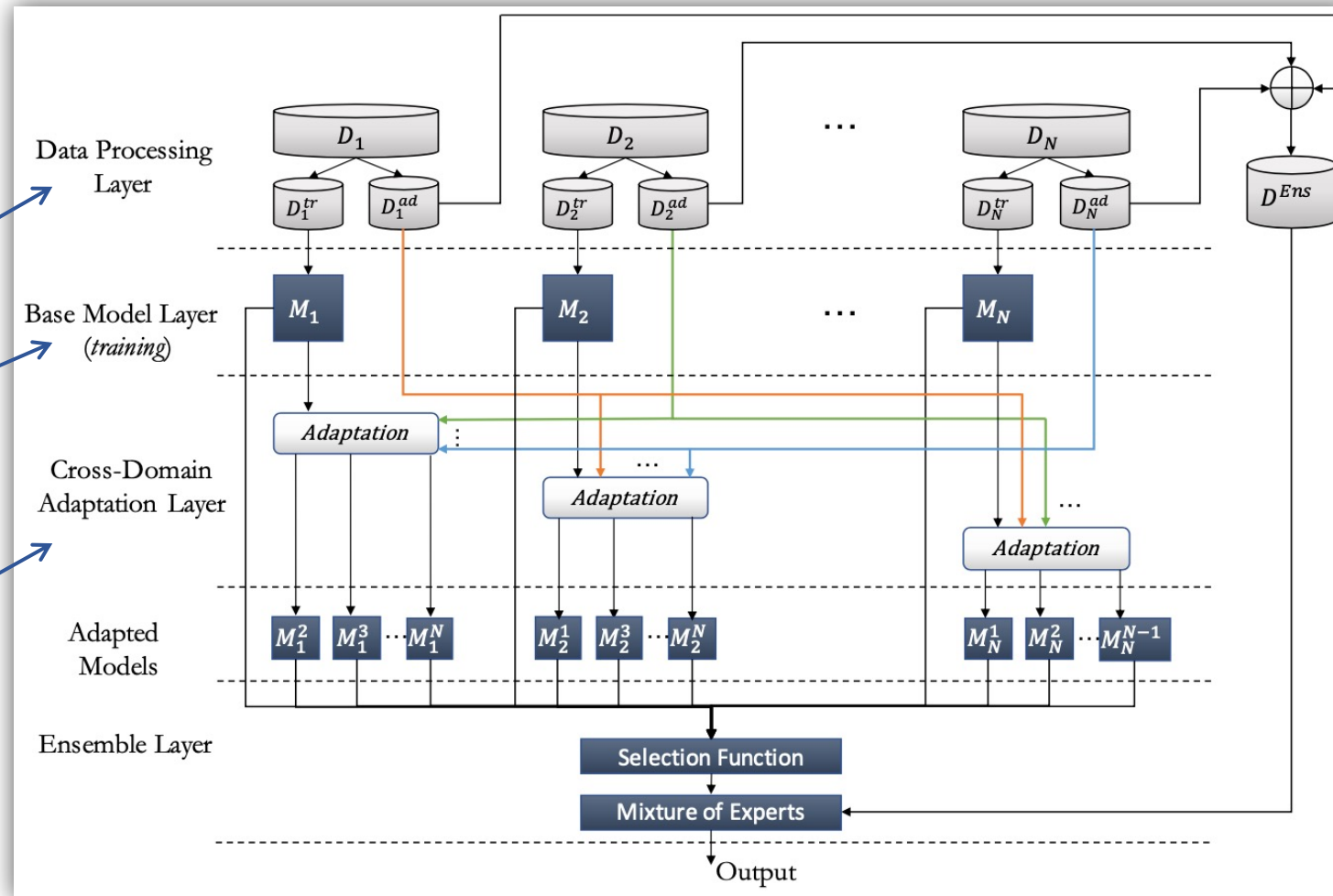


Proposed Solution: MoDA - FND

1 Split datasets

2 Train a base specialized fake news detector

3 Adapt the model



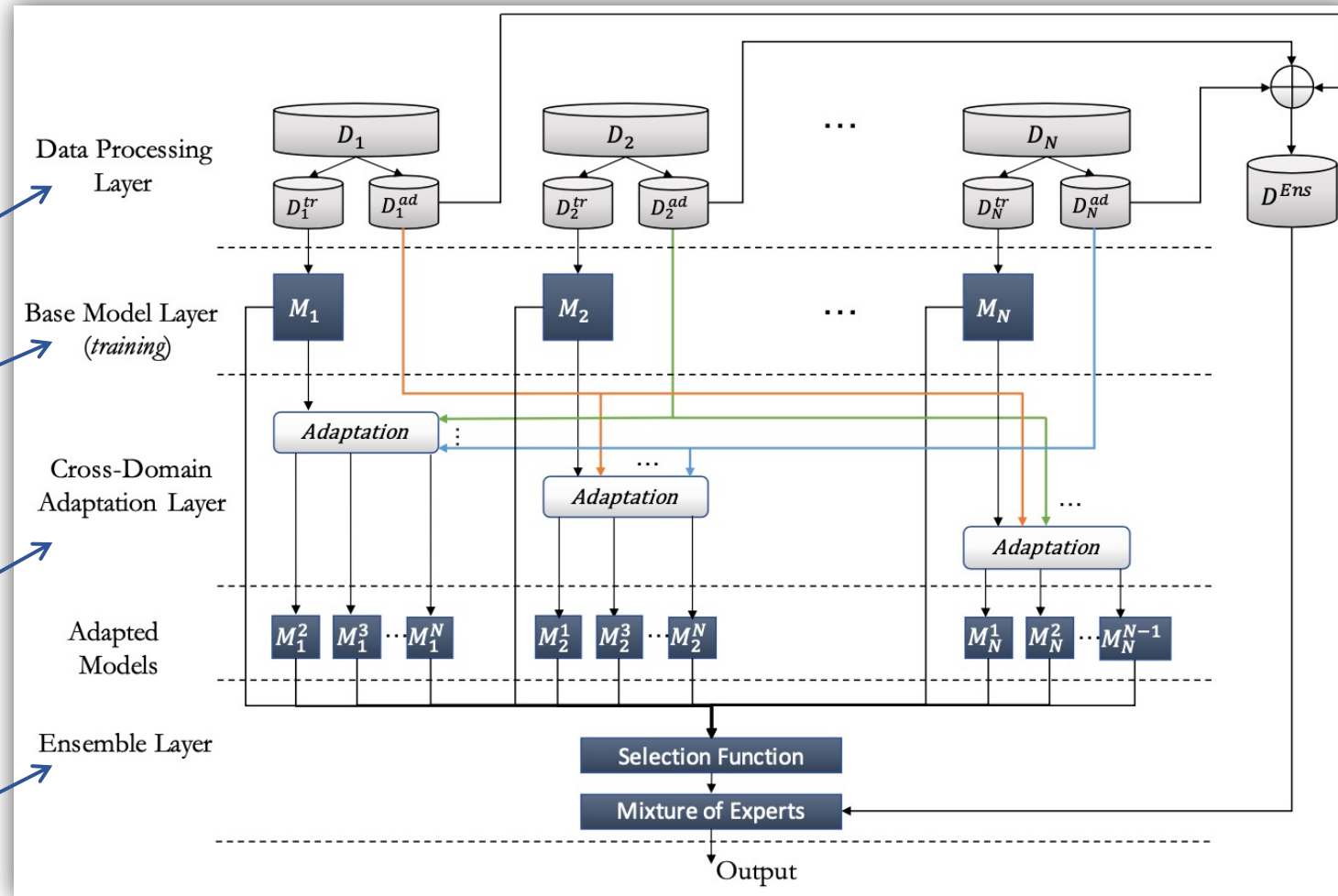
Proposed Solution: MoDA - FND

1 Split datasets

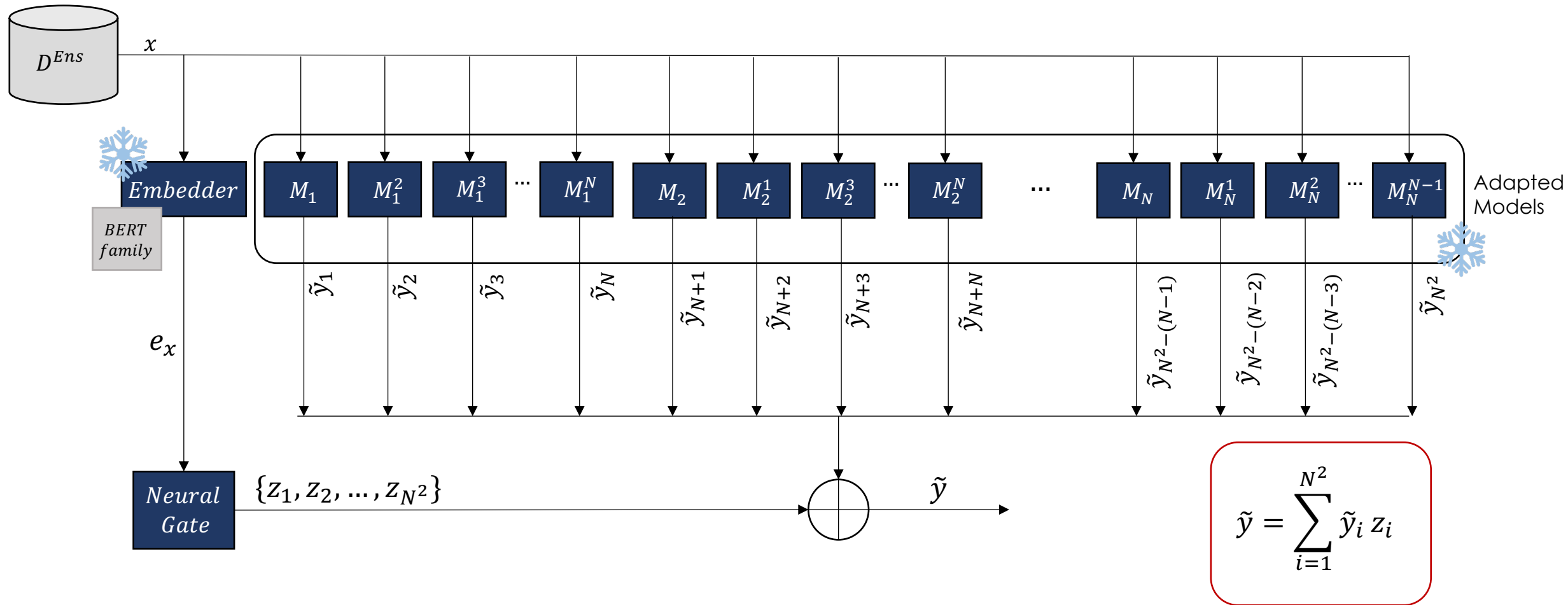
2 Train a base specialized fake news detector

3 Adapt the model

4 Train the ensemble model

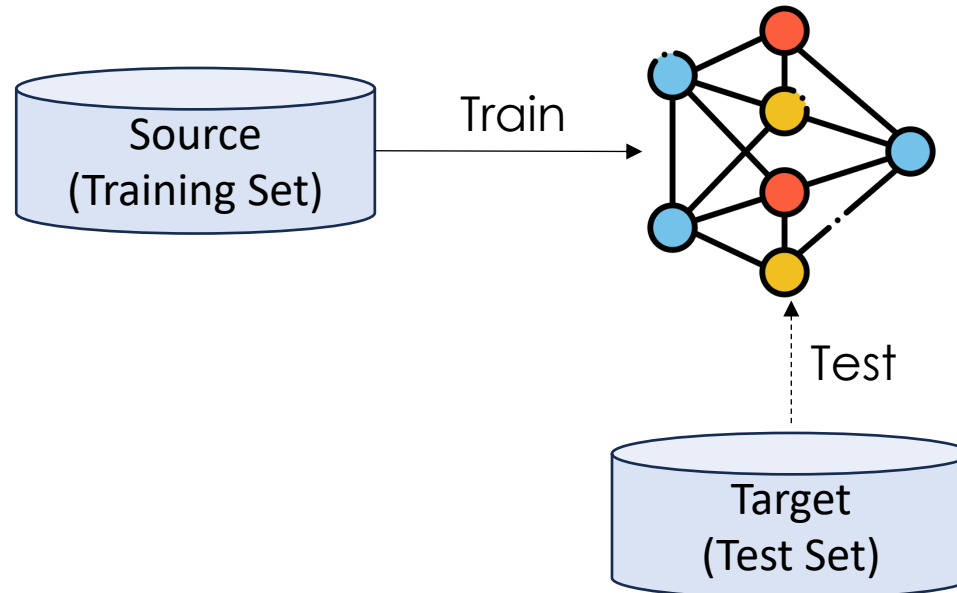


Proposed Solution: Mixture of Experts



Experimental Assessment

Dataset	#Articles	#Real	#Fake	Vocabulary size	#Words per article			
					Avg.	Median	Q1	Q3
GossipCop	4,736	2,117	2,619	167,759	581	348	180	551
PolitiFact	818	370	448	69,796	1,470	336	108	882
MD	6,296	3,171	3,125	216,169	781	601	294	1,026
CoAID	4,439	4,023	416	15,738	56	74	14	80
Liar	12,685	9,183	3,502	24,549	18	17	12	22



Experimental Assessment

RQ1. How does *MoDA – FND* compare to SOTA approaches?

	GossipCop	Politifact	MD	CoAID	Liar
CrossFND	0.4374	0.3717	0.5415	0.2259	0.4880
EANN	0.5122	0.3477	0.5503	0.5303	0.4199
MDA-WS	0.3089	0.3539	0.3317	0.4754	0.2174
<i>MoDA – FND</i>	0.4857	0.4817	0.6474	0.5660	0.5045

Experimental Assessment

RQ2. How does the MoE-based ensemble strategy compare to other ensemble strategies?

	GossipCop	Politifact	MD	CoAID	Liar
Max-Probability	0.4939	0.2896	0.4745	0.2548	0.5215
Stacking	0.5042	0.4201	0.4117	0.4423	0.5079
Stacking - HP	0.4990	0.4026	0.4980	0.5251	0.5077
<i>MoDA – FND</i>	0.4857	0.4817	0.6474	0.5660	0.5045

Conclusion

- *MoDA-FND, a hierarchical deep ensemble approach for detecting fake news*
 - Recognize fake news across different domains
 - Exploiting one learning model able to generalize through the domains
- As future work, we are interested in enhancing the comprehensiveness and effectiveness of *MoDA-FND* by integrating additional sources of information



Thank you for your attention!
Questions?