Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

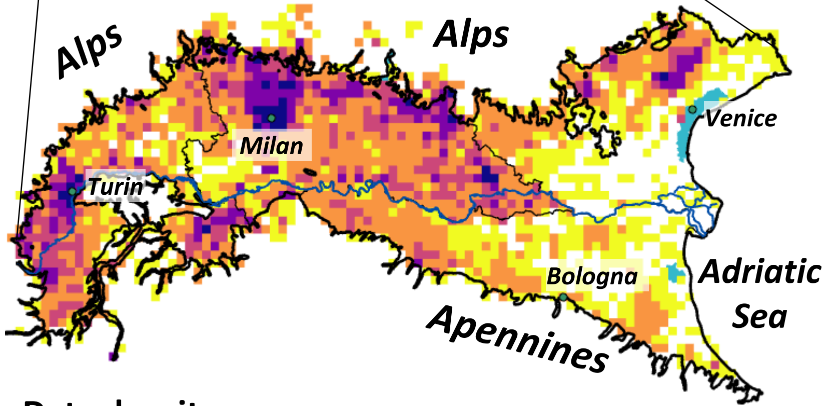
Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> **2 min** **FULL**

Study area

Po River Alluvial Plain

45’700 km^2



Data density (bh/km^2)

- ☐ No Data
- 🟨 < 0.2
- 🟧 < 1.0
- 🟥 < 2.0
- 🟪 < 5.0
- 🟦 < 20

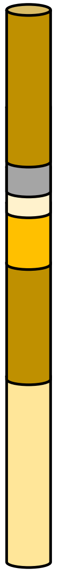
LSTM Classification Net

86,611 lithological descriptions

(70% training - 30% validation)

→ **39,265 boreholes**
387,297 descriptive intervals

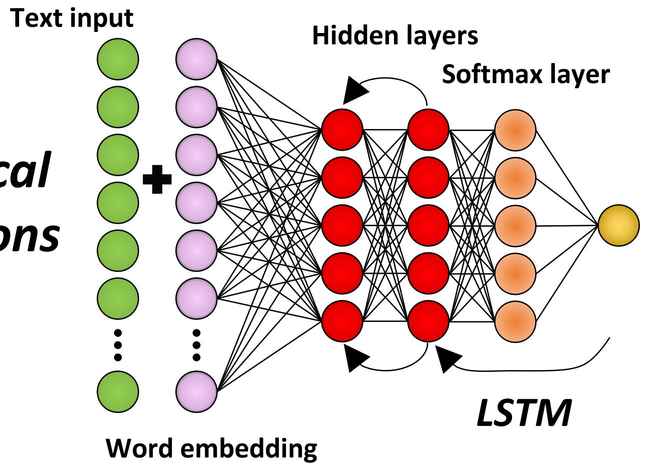
Borehole logs

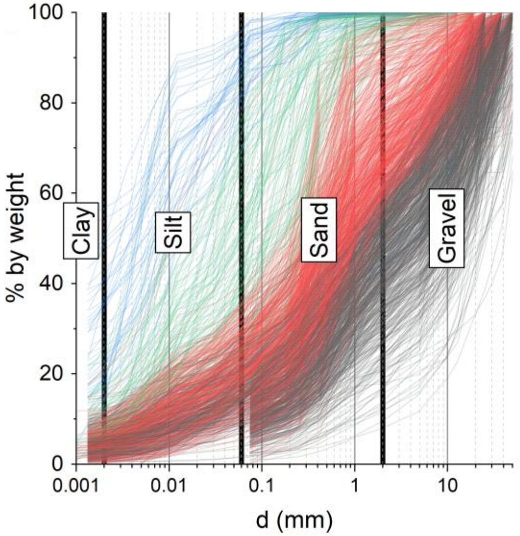


Lithological descriptions

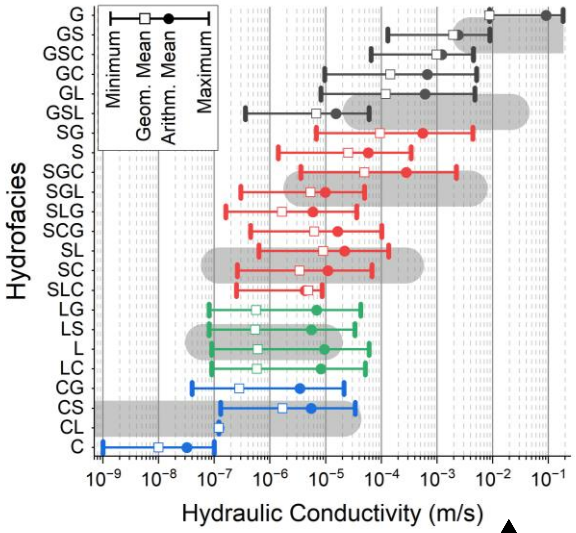
Grain size analysis

Deep learning classification





Classification



Parametrization

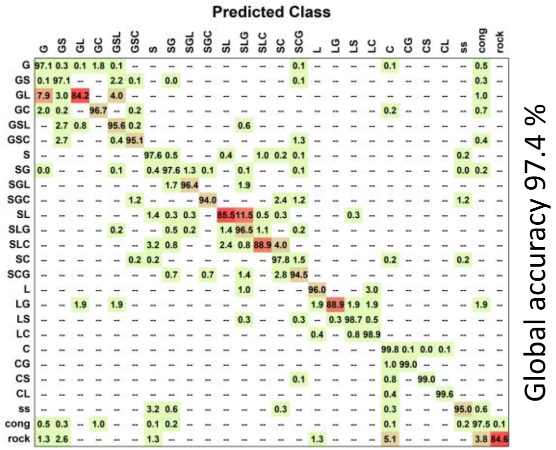
Empirical equations from grain size distributions

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

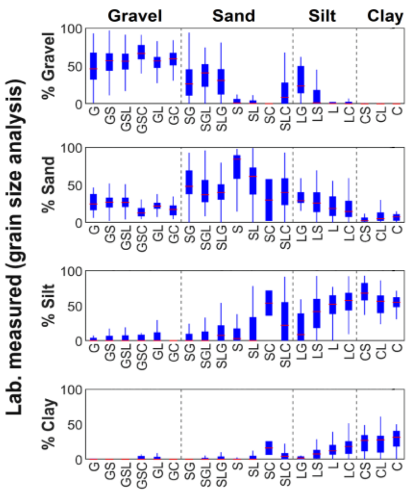
Navigate -> **2 min** **FULL**

Confusion matrix



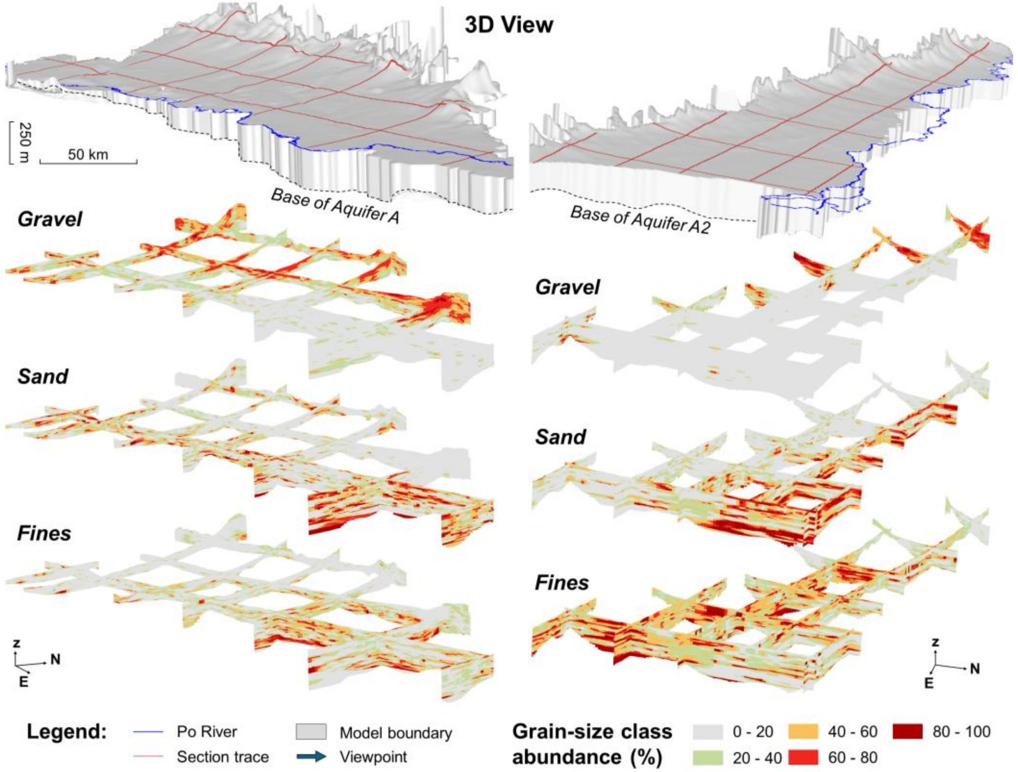
	G	GS	GL	GC	GSL	GSC	S	SG	SGL	SGC	SL	SLG	SLC	SC	SCG	L	LG	LS	LC	C	CG	CS	CL	ss	cong	rock
G	97.1	0.3	0.1	1.8	0.1	--	--	--	--	--	--	--	--	--	0.1	--	--	--	--	0.1	--	--	--	--	0.5	--
GS	0.1	97.1	--	--	2.2	0.1	--	0.0	--	--	--	--	--	--	0.1	--	--	--	--	--	--	--	--	--	0.3	--
GL	7.9	3.0	84.2	--	4.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1.0	--
GC	2.0	0.2	--	96.7	--	0.2	--	--	--	--	--	--	--	--	--	--	--	--	--	0.2	--	--	--	--	0.7	--
GSL	--	2.7	0.8	--	95.6	0.2	--	--	--	--	--	0.6	--	--	--	--	--	--	--	--	--	--	--	--	--	--
GSC	--	2.7	--	--	0.4	95.1	--	--	--	--	--	--	--	--	1.3	--	--	--	--	--	--	--	--	--	0.4	--
S	--	--	--	--	--	--	97.6	0.5	--	--	0.4	--	1.0	0.2	0.1	--	--	--	--	--	--	--	--	0.2	--	--
SG	0.0	--	--	--	0.1	--	0.4	97.6	1.3	0.1	--	0.1	--	--	0.1	--	--	--	--	--	--	--	--	0.0	0.2	--
SGL	--	--	--	--	--	--	--	1.7	96.4	--	--	1.9	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SGC	--	--	--	--	--	1.2	--	--	--	94.0	--	--	--	2.4	1.2	--	--	--	--	--	--	--	--	1.2	--	--
SL	--	--	--	--	--	--	1.4	0.3	0.3	--	85.5	11.5	0.5	0.3	--	--	--	0.3	--	--	--	--	--	--	--	--
SLG	--	--	--	--	0.2	--	--	0.5	0.2	--	1.4	96.5	1.1	--	0.2	--	--	--	--	--	--	--	--	--	--	--
SLC	--	--	--	--	--	--	3.2	0.8	--	--	2.4	0.8	88.9	4.0	--	--	--	--	--	--	--	--	--	--	--	--
SC	--	--	--	--	--	0.2	0.2	--	--	--	--	--	--	97.8	1.5	--	--	--	--	0.2	--	--	--	0.2	--	--
SCG	--	--	--	--	--	--	--	0.7	--	0.7	--	1.4	--	2.8	94.5	--	--	--	--	--	--	--	--	--	--	--
L	--	--	--	--	--	--	--	--	--	--	--	1.0	--	--	--	96.0	--	--	3.0	--	--	--	--	--	--	--
LG	--	--	1.9	--	1.9	--	--	--	--	--	--	--	--	--	--	1.9	88.9	1.9	1.9	--	--	--	--	--	1.9	--
LS	--	--	--	--	--	--	--	--	--	--	--	0.3	--	--	0.3	--	0.3	98.7	0.5	--	--	--	--	--	--	--
LC	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.4	--	0.8	98.9	--	--	--	--	--	--	--
C	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	99.8	0.1	0.0	0.1	--	--	--
CG	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1.0	99.0	--	--	--	--	--
CS	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.1	--	--	--	--	0.8	--	99.0	--	--	--	--
CL	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.4	--	--	99.6	--	--	--
ss	--	--	--	--	--	--	3.2	0.6	--	--	--	--	--	0.3	--	--	--	--	--	0.3	--	--	--	95.0	0.6	--
cong	0.5	0.3	--	1.0	--	--	0.1	0.2	--	--	--	--	--	--	--	--	--	--	--	0.1	--	--	--	0.2	97.5	0.1
rock	1.3	2.6	--	--	--	--	1.3	--	--	--	--	--	--	--	--	1.3	--	--	--	5.1	--	--	--	--	3.8	84.6

Predicted grain size



Spatialization

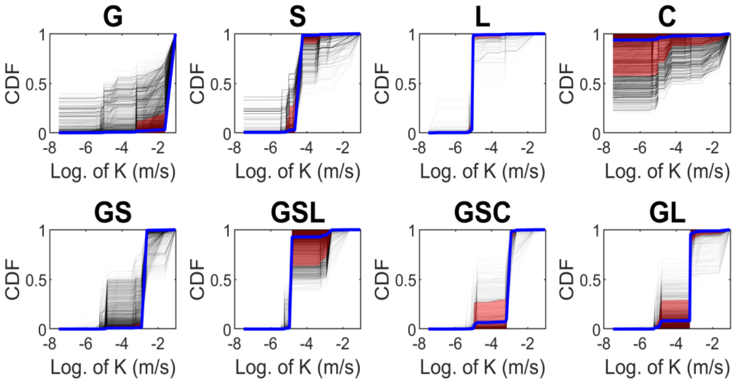
3D interpolation of grain size and permeability values attributed to borehole logs

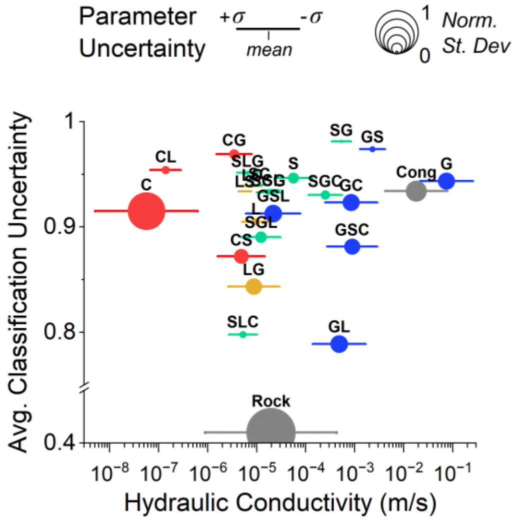


Increased data density for modeling from 0.3 to 8.7 data/km^2

Classification uncertainty

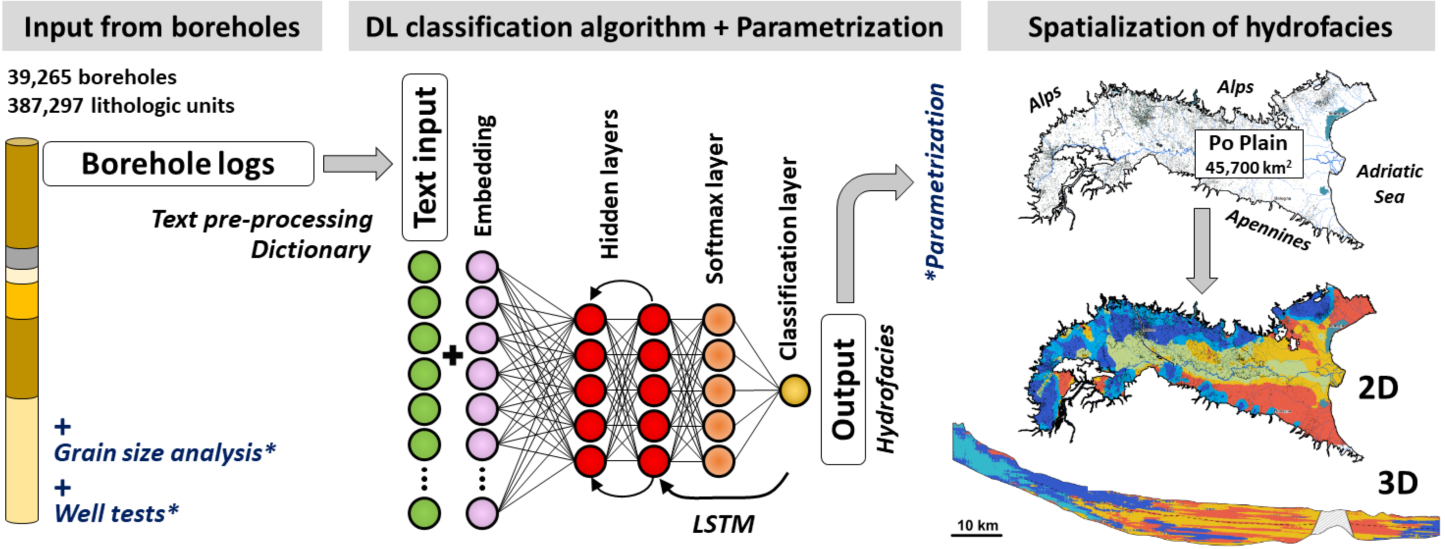
Probabilistic parametrization from the DL classification uncertainty





Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157



- Deep learning text classification of borehole geological/stratigraphic descriptions
- Recurrent neural network (RNN) with long short-term memory (LSTM) and word embedding

- Hydrofacies classification of litho-textural logs for the Po River alluvial plain
- Hydraulic conductivity associated to hydrofacies classes from the sediment grain size

- Hydrofacies models generated from new semi-quantitative hydrogeological data

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate ↓

Introduction

Study area

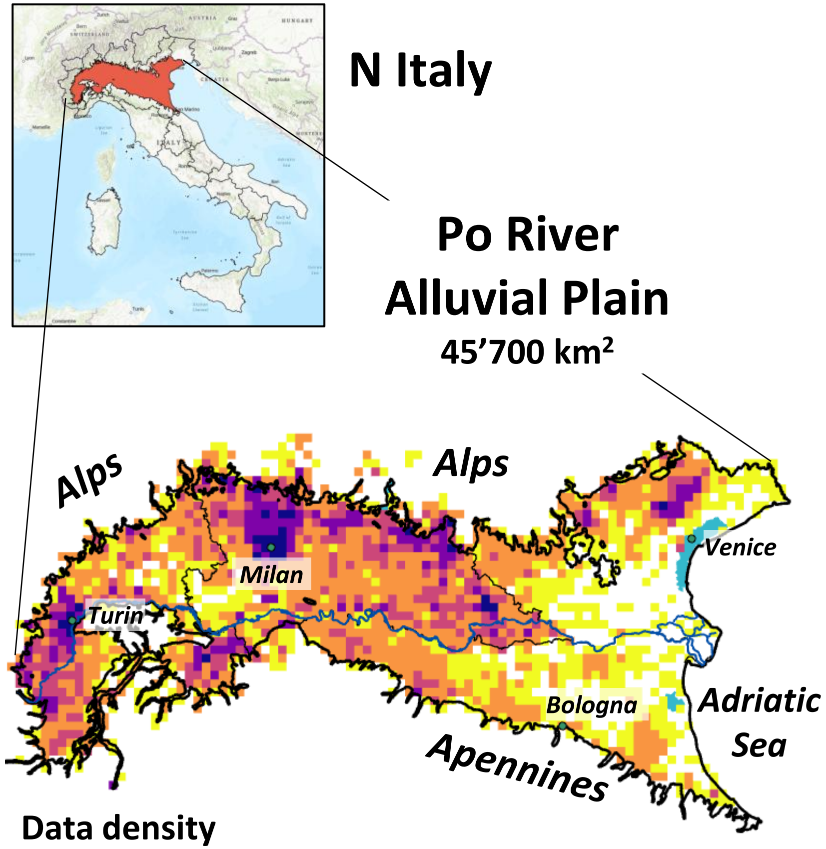
Methods

Results

Classification

Spatialization

Conclusions



N Italy

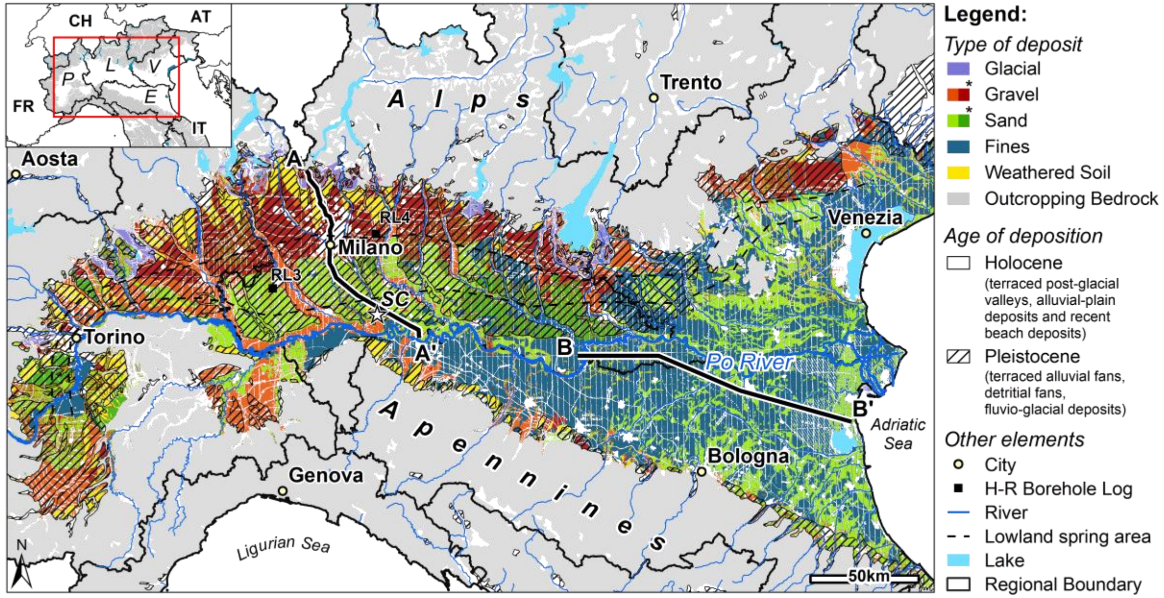
Po River Alluvial Plain
45'700 km^2

Data density (bh/km^2)

- ☐ No Data
- 🟨 < 0.2
- 🟧 < 1.0
- 🟥 < 2.0
- 🟪 < 5.0
- 🟦 < 20

39,265 boreholes
387,297 descriptive intervals
(70% training - 30% validation)

Geological setting



Neogenic-Quaternary siliciclastic sequence that blankets the external fronts of mountain chains, composed of syntectonic and recent alluvial sediments from the Po River

Hydrostratigraphic units forming regional-scale aquifer groups

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓

Introduction

Study area

Methods

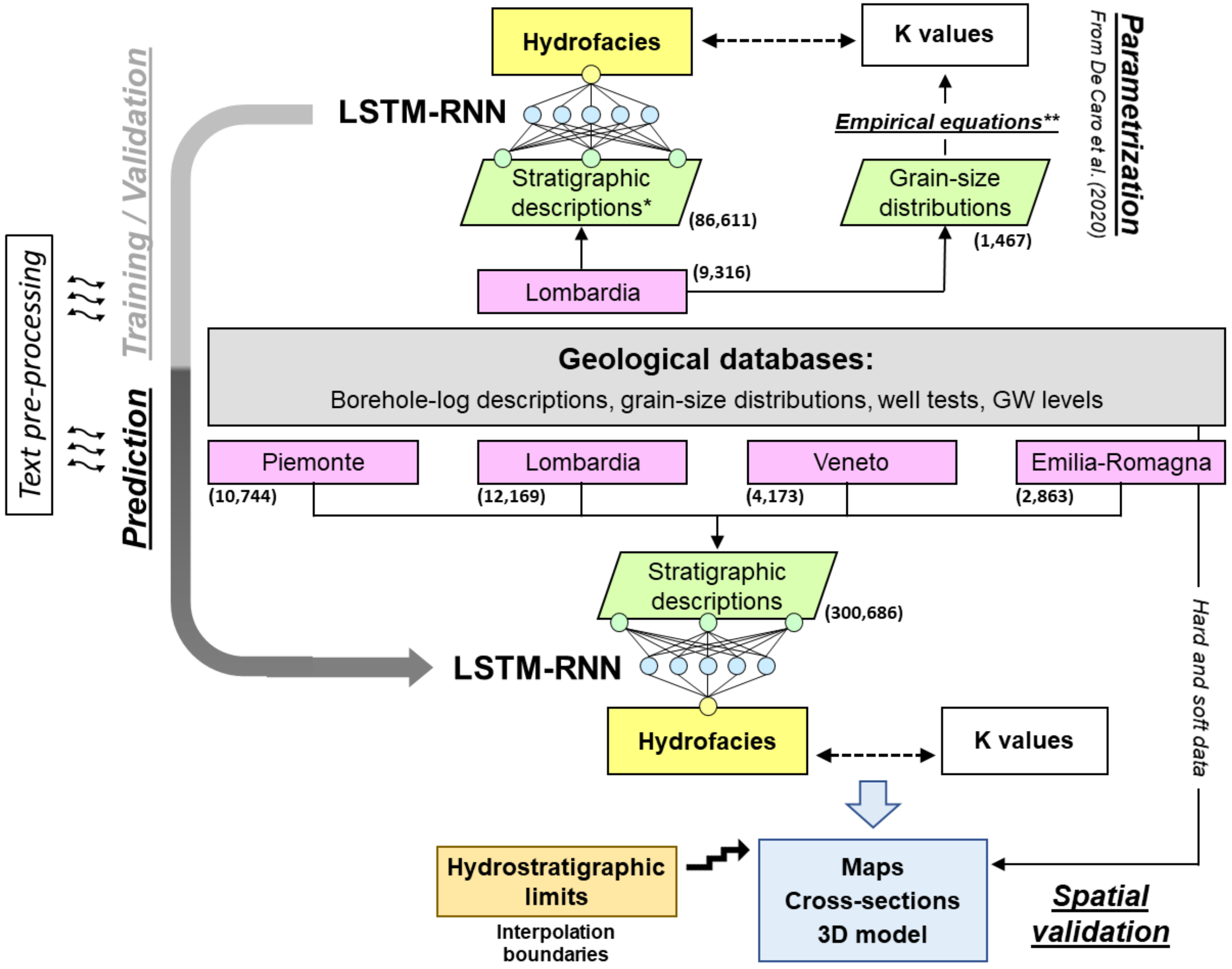
Results

Classification

Spatialization

Conclusions

WORKFLOW



Hydrofacies ← - - → K values

LSTM-RNN

Empirical equations**

Stratigraphic descriptions* (86,611)

Grain-size distributions (1,467)

Lombardia (9,316)

Parametrization

From De Caro et al. (2020)

Text pre-processing

Training / Validation

Prediction

Geological databases:

Borehole-log descriptions, grain-size distributions, well tests, GW levels

Piemonte (10,744)

Lombardia (12,169)

Veneto (4,173)

Emilia-Romagna (2,863)

Stratigraphic descriptions (300,686)

LSTM-RNN

Hydrofacies ← - - → K values

Hard and soft data

Hydrostratigraphic limits

Interpolation boundaries

Maps
Cross-sections
3D model

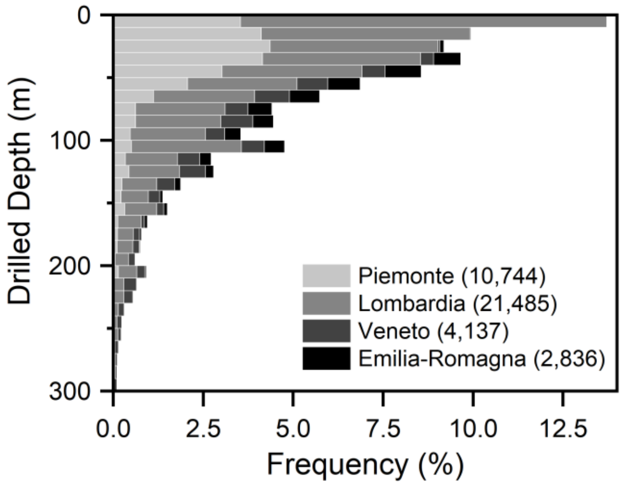
Spatial validation

LSTM Classification Net

86,611 lithological descriptions

(70% training - 30% validation)

Borehole length



39,265 boreholes

387,297 descriptive intervals

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓

Introduction

Study area

Methods

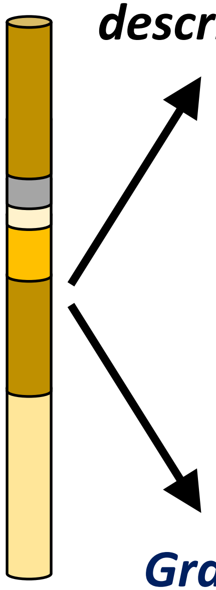
Results

Classification

Spatialization

Conclusions

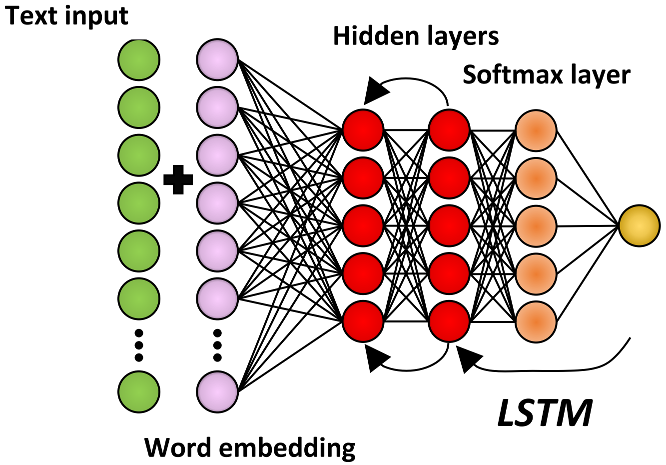
Borehole logs

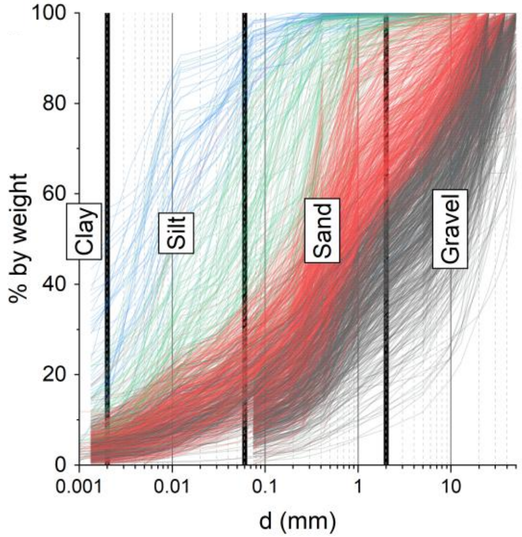


Lithological descriptions

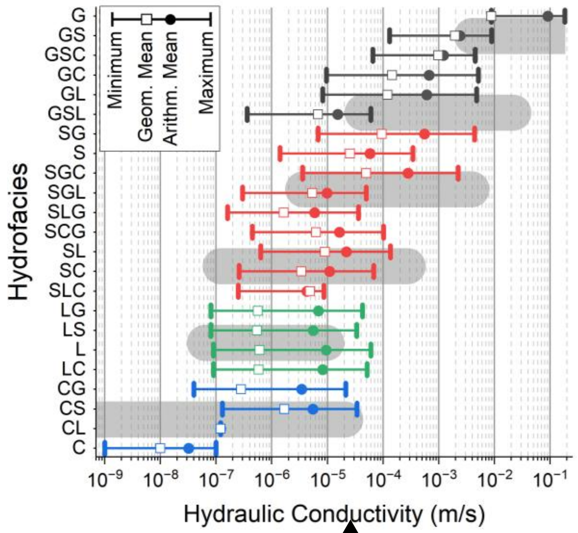
Grain size analysis

Deep learning classification





Classification



Parametrization

Empirical equations from grain size distributions

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓
Introduction
Study area
Methods
Results
Classification
Spatialization
Conclusions

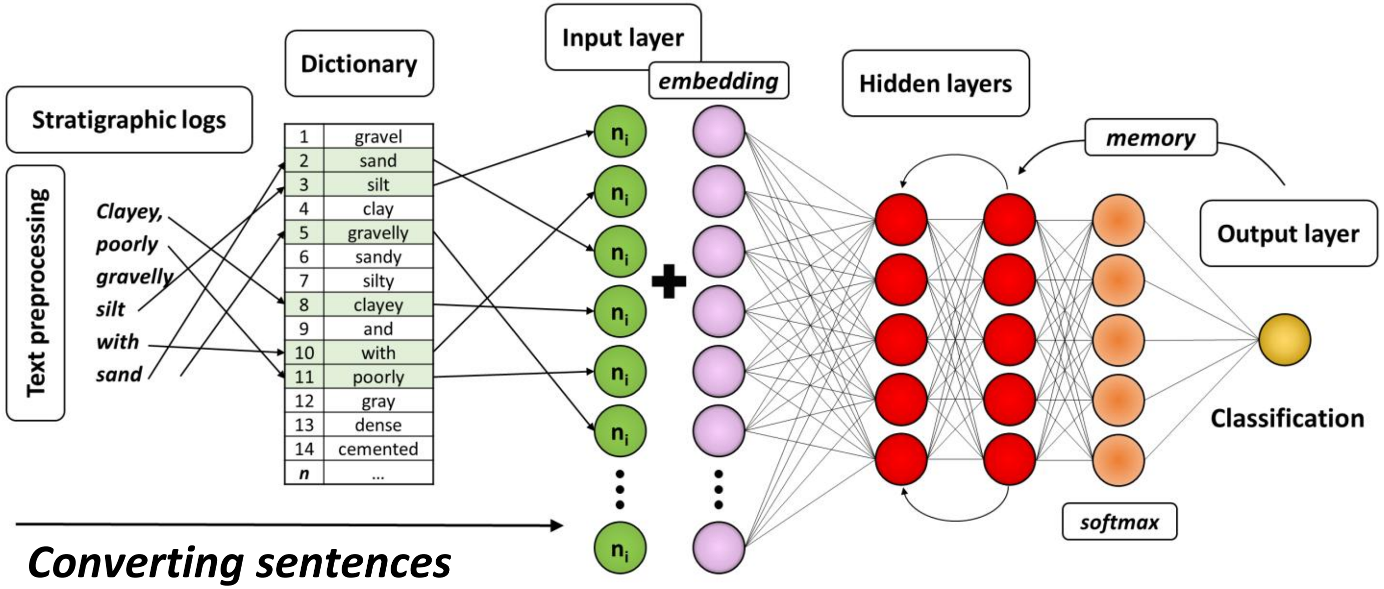
Text pre-processing

- *Remove punctuation*
- *Remove uppercase*
- *Mistakes identification*

Dictionary
(In Italian)

≈ 3000 terms

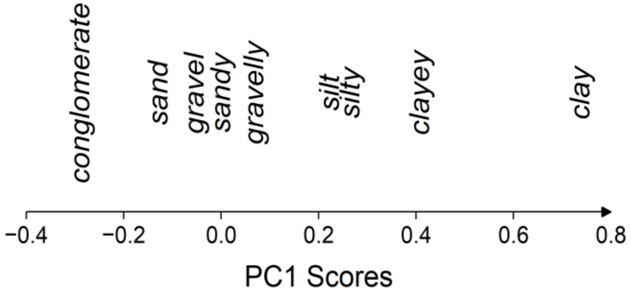
Features of the geological text deep learning classification algorithm



1	gravel
2	sand
3	silt
4	clay
5	gravelly
6	sandy
7	silty
8	clayey
9	and
10	with
11	poorly
12	gray
13	dense
14	cemented
n	...

Converting sentences into numerical sequences

Classification learning process RNN + LSTM



The word embedding vectors for the most frequent litho-textural terms!

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min | FULL

Navigate ↓

Introduction

Study area

Methods

Results

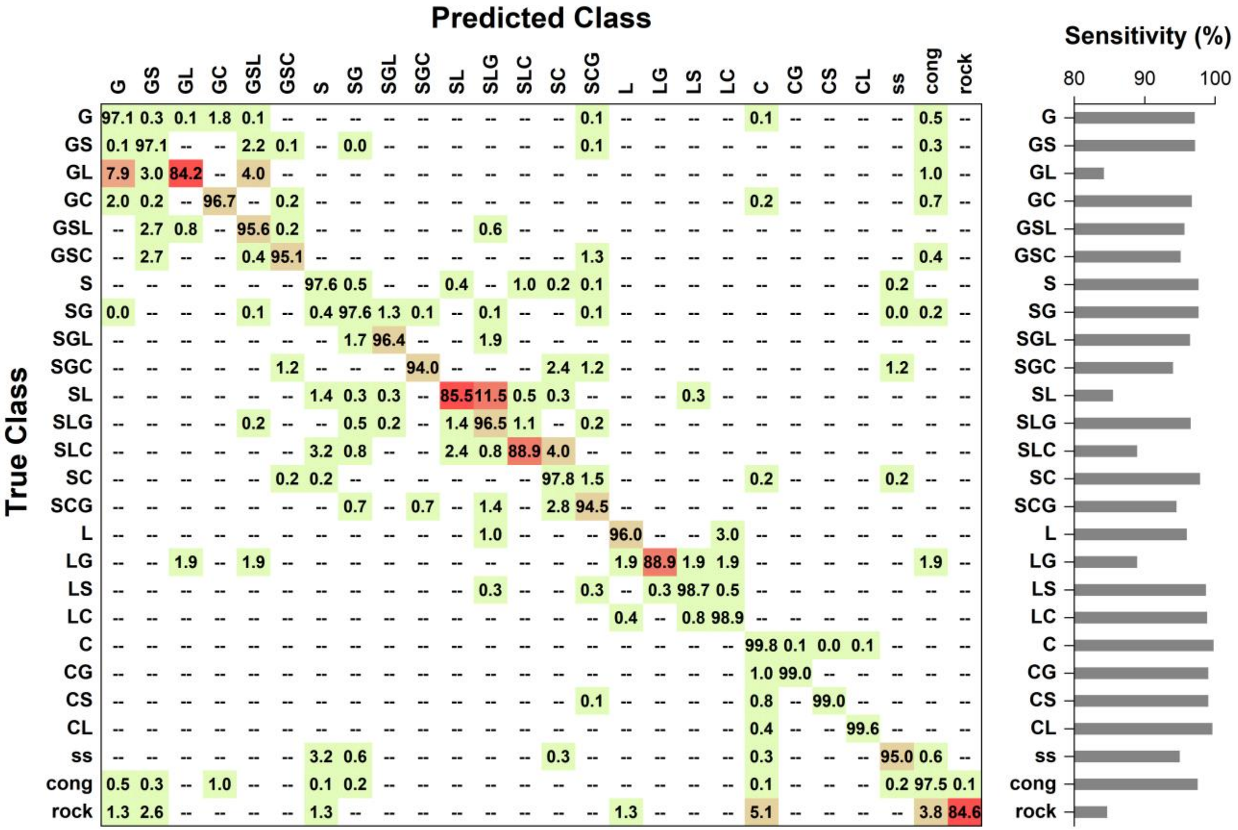
Classification

Spatialization

Conclusions

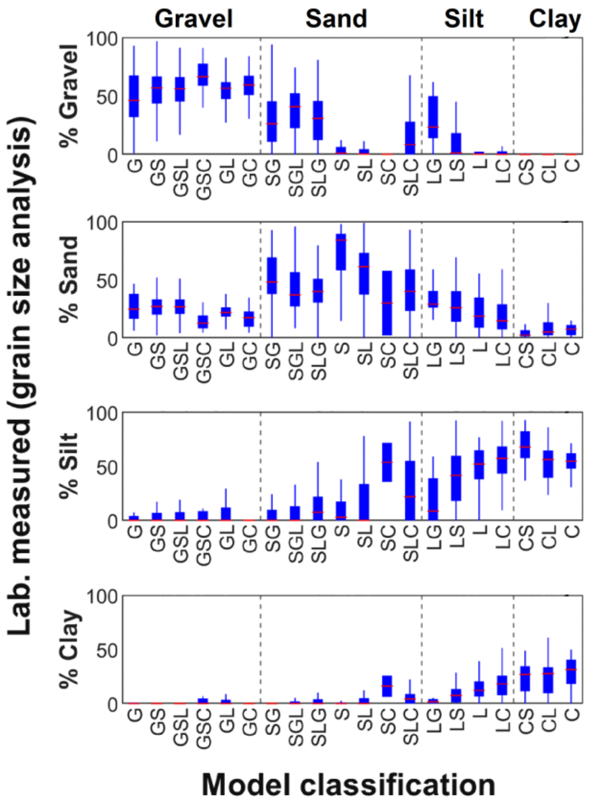
Classification confusion matrix

global validation accuracy (97.4%)



	G	GS	GL	GC	GSL	GSC	S	SG	SGL	SGC	SL	SLG	SLC	SC	SCG	L	LG	LS	LC	C	CG	CS	CL	ss	cong	rock
G	97.1	0.3	0.1	1.8	0.1	--	--	--	--	--	--	--	--	--	0.1	--	--	--	--	0.1	--	--	--	--	0.5	--
GS	0.1	97.1	--	--	2.2	0.1	--	0.0	--	--	--	--	--	--	0.1	--	--	--	--	--	--	--	--	--	0.3	--
GL	7.9	3.0	84.2	--	4.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1.0	--
GC	2.0	0.2	--	96.7	--	0.2	--	--	--	--	--	--	--	--	--	--	--	--	--	0.2	--	--	--	--	0.7	--
GSL	--	2.7	0.8	--	95.6	0.2	--	--	--	--	--	0.6	--	--	--	--	--	--	--	--	--	--	--	--	--	--
GSC	--	2.7	--	--	0.4	95.1	--	--	--	--	--	--	--	--	1.3	--	--	--	--	--	--	--	--	--	0.4	--
S	--	--	--	--	--	--	97.6	0.5	--	--	0.4	--	1.0	0.2	0.1	--	--	--	--	--	--	--	--	0.2	--	--
SG	0.0	--	--	--	0.1	--	0.4	97.6	1.3	0.1	--	0.1	--	--	0.1	--	--	--	--	--	--	--	--	0.0	0.2	--
SGL	--	--	--	--	--	--	--	1.7	96.4	--	--	1.9	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SGC	--	--	--	--	--	1.2	--	--	--	94.0	--	--	--	2.4	1.2	--	--	--	--	--	--	--	--	1.2	--	--
SL	--	--	--	--	--	--	1.4	0.3	0.3	--	85.5	11.5	0.5	0.3	--	--	--	0.3	--	--	--	--	--	--	--	--
SLG	--	--	--	--	0.2	--	--	0.5	0.2	--	1.4	96.5	1.1	--	0.2	--	--	--	--	--	--	--	--	--	--	--
SLC	--	--	--	--	--	--	3.2	0.8	--	--	2.4	0.8	88.9	4.0	--	--	--	--	--	--	--	--	--	--	--	--
SC	--	--	--	--	--	0.2	0.2	--	--	--	--	--	--	97.8	1.5	--	--	--	--	0.2	--	--	--	0.2	--	--
SCG	--	--	--	--	--	--	--	0.7	--	0.7	--	1.4	--	2.8	94.5	--	--	--	--	--	--	--	--	--	--	--
L	--	--	--	--	--	--	--	--	--	--	--	1.0	--	--	--	96.0	--	--	3.0	--	--	--	--	--	--	--
LG	--	--	1.9	--	1.9	--	--	--	--	--	--	--	--	--	--	1.9	88.9	1.9	1.9	--	--	--	--	--	1.9	--
LS	--	--	--	--	--	--	--	--	--	--	--	0.3	--	--	0.3	--	0.3	98.7	0.5	--	--	--	--	--	--	--
LC	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.4	--	0.8	98.9	--	--	--	--	--	--	--
C	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	99.8	0.1	0.0	0.1	--	--	--
CG	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1.0	99.0	--	--	--	--	--
CS	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.1	--	--	--	--	0.8	--	99.0	--	--	--	--
CL	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.4	--	--	99.6	--	--	--
ss	--	--	--	--	--	--	3.2	0.6	--	--	--	--	--	0.3	--	--	--	--	--	0.3	--	--	--	95.0	0.6	--
cong	0.5	0.3	--	1.0	--	--	0.1	0.2	--	--	--	--	--	--	--	--	--	--	--	0.1	--	--	--	0.2	97.5	0.1
rock	1.3	2.6	--	--	--	--	1.3	--	--	--	--	--	--	--	--	1.3	--	--	--	5.1	--	--	--	--	3.8	84.6

Predicted vs measured grain size



G = gravel, S = sand, L = silt, C = clay, ss = sandstone, cong = conglomerate, and rock = generic rock type.

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> **2 min** **FULL**

Navigate ↓

Introduction

Study area

Methods

Results

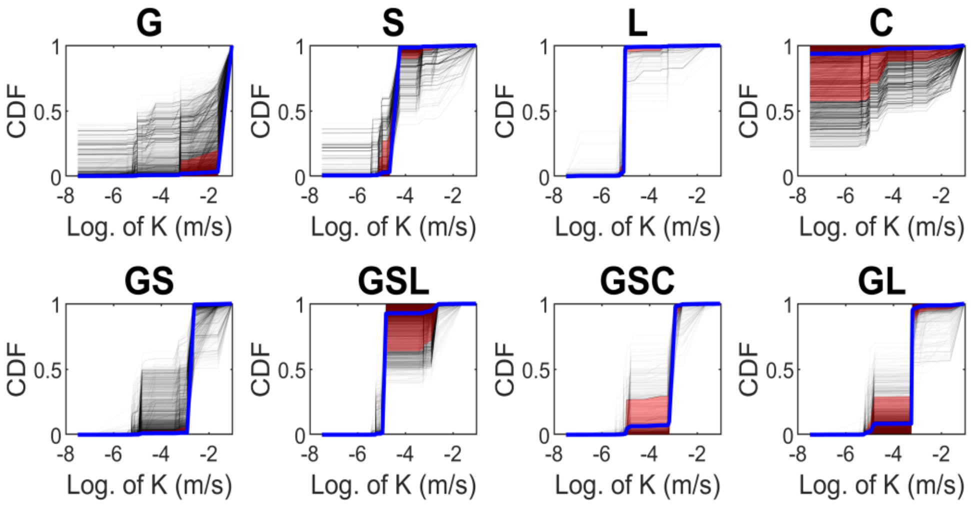
Classification

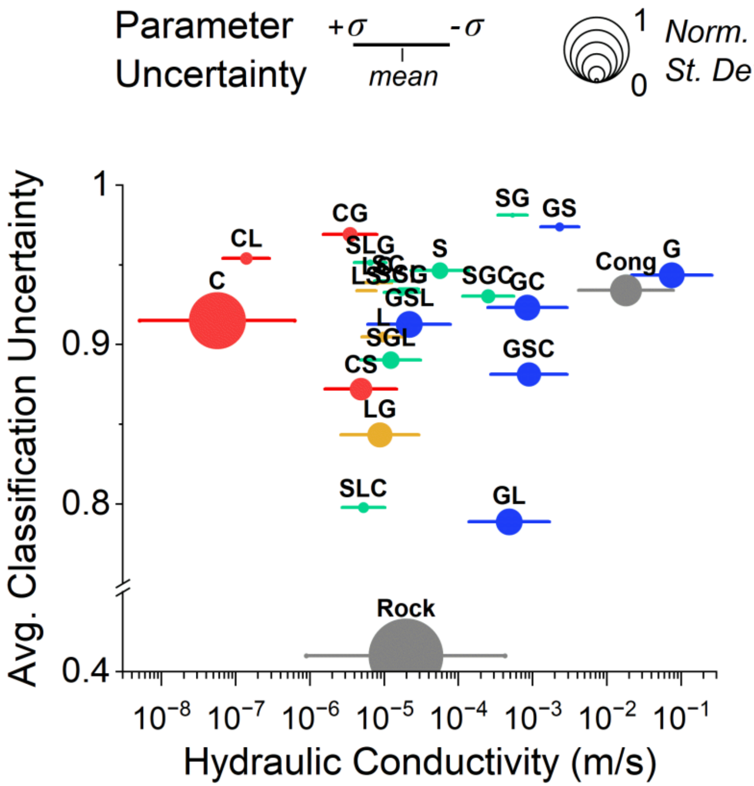
Spatialization

Conclusions

Classification uncertainty

Probabilistic parametrization from the DL classification uncertainty (***softmax*** **layer**)





The resulting uncertainty in the associated parameter (K) can vary significantly, depending on the classifier's confidence in assigning alternative classes.

- Some classes are confidently classified and have a low parameter uncertainty because the few indecisions involve classes with similar values
- For other classes, despite being relatively confident, the parameter uncertainty is higher due to the markedly different K values of potential alternative classifications

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓

Introduction

Study area

Methods

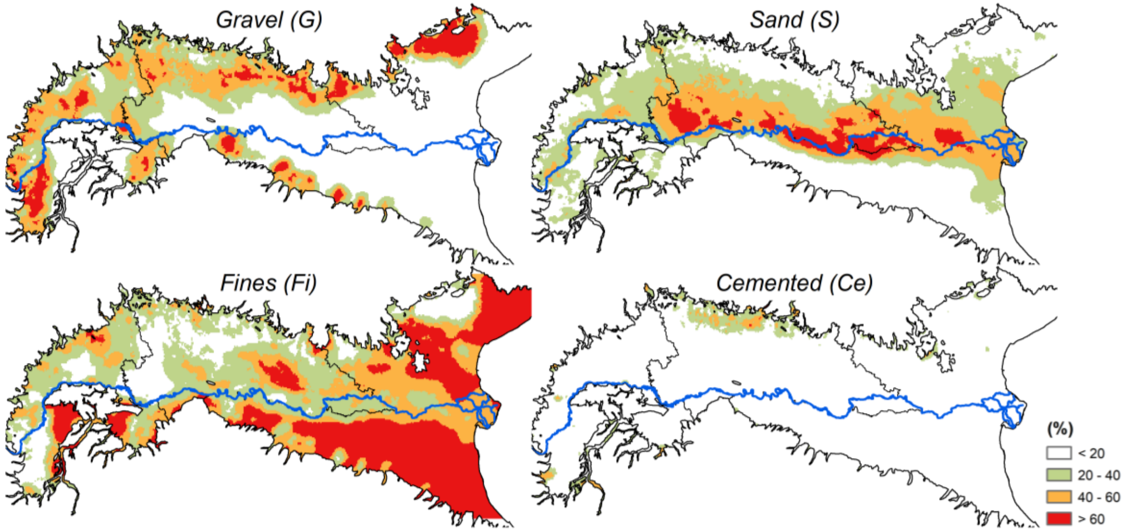
Results

Classification

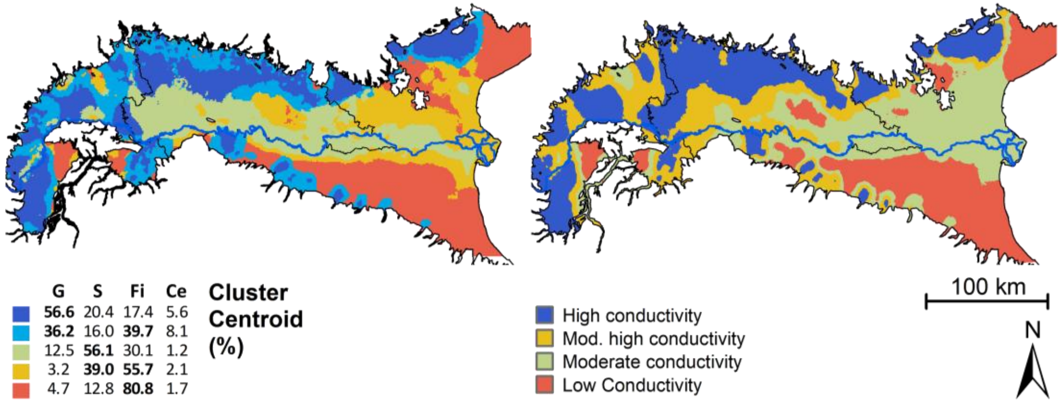
Spatialization

Conclusions

Grain size maps (50m depth)



Hydrogeological maps (50m depth)

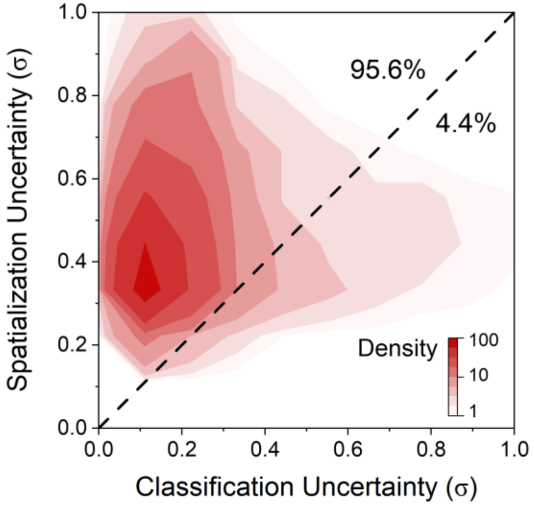


	G	S	Fi	Ce
🟦	**56.6**	20.4	17.4	5.6
🟦	**36.2**	16.0	**39.7**	8.1
🟩	12.5	**56.1**	30.1	1.2
🟨	3.2	**39.0**	**55.7**	2.1
🟥	4.7	12.8	**80.8**	1.7

Cluster Centroid (%)

🟦 High conductivity
🟨 Mod. high conductivity
🟩 Moderate conductivity
🟥 Low Conductivity

Spatialization vs Classification Uncertainty



Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓

Introduction

Study area

Methods

Results

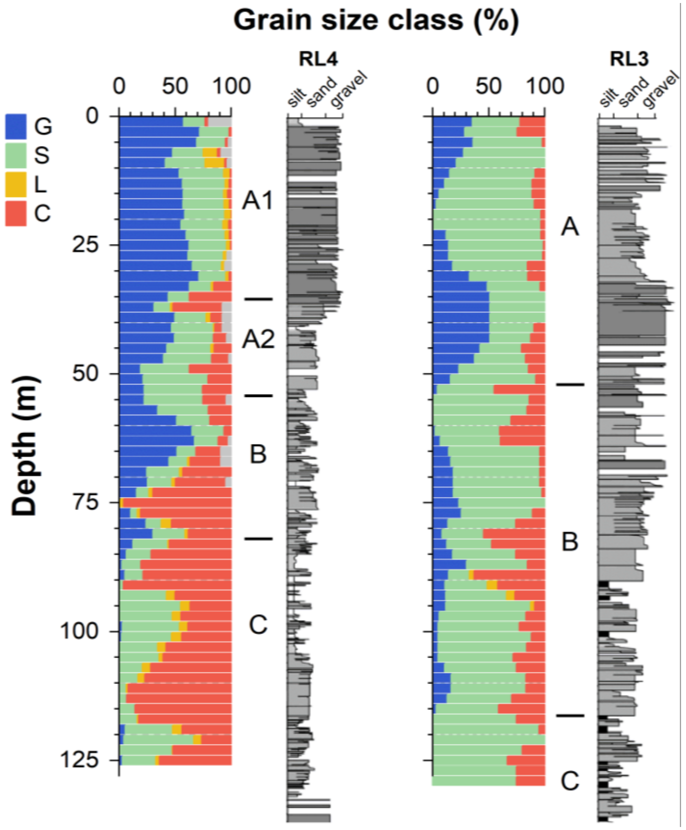
Classification

Spatialization

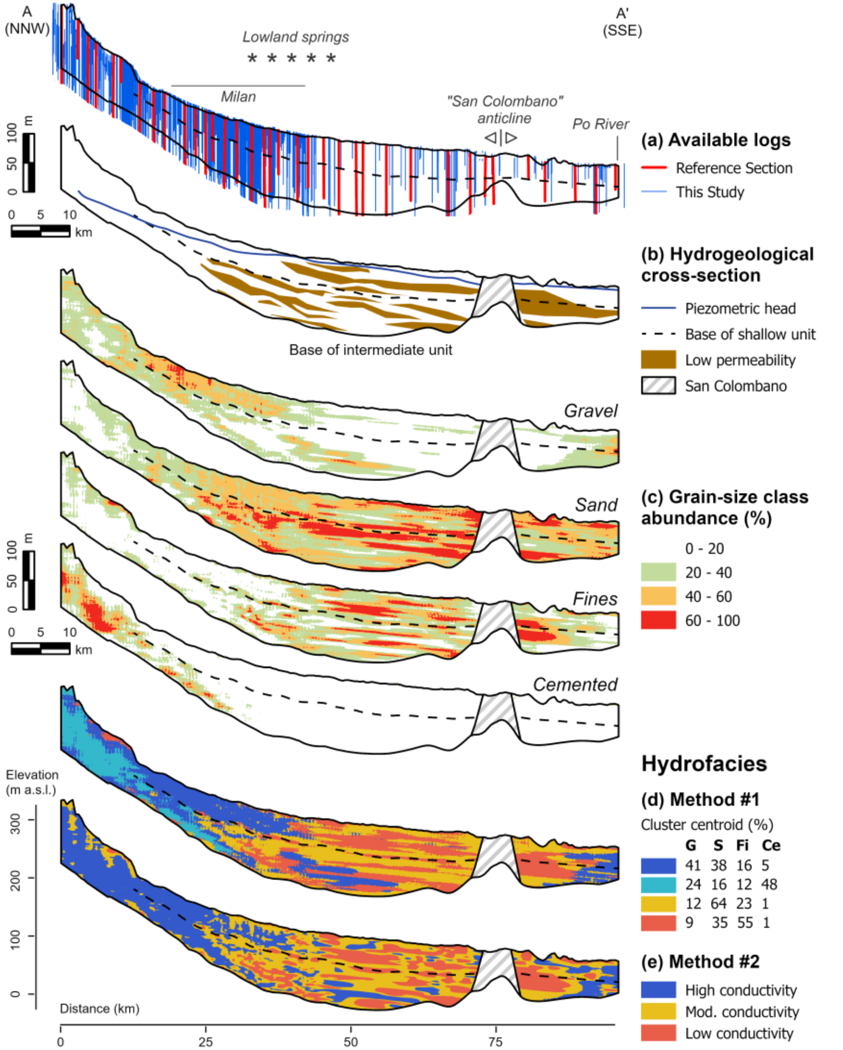
Conclusions

1D comparison of grain size logs:

- Identification of main aquifers



Creation of hydrogeological cross-sections



Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – *Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157*

Navigate -> **2 min** **FULL**

Navigate ↓

Introduction

Study area

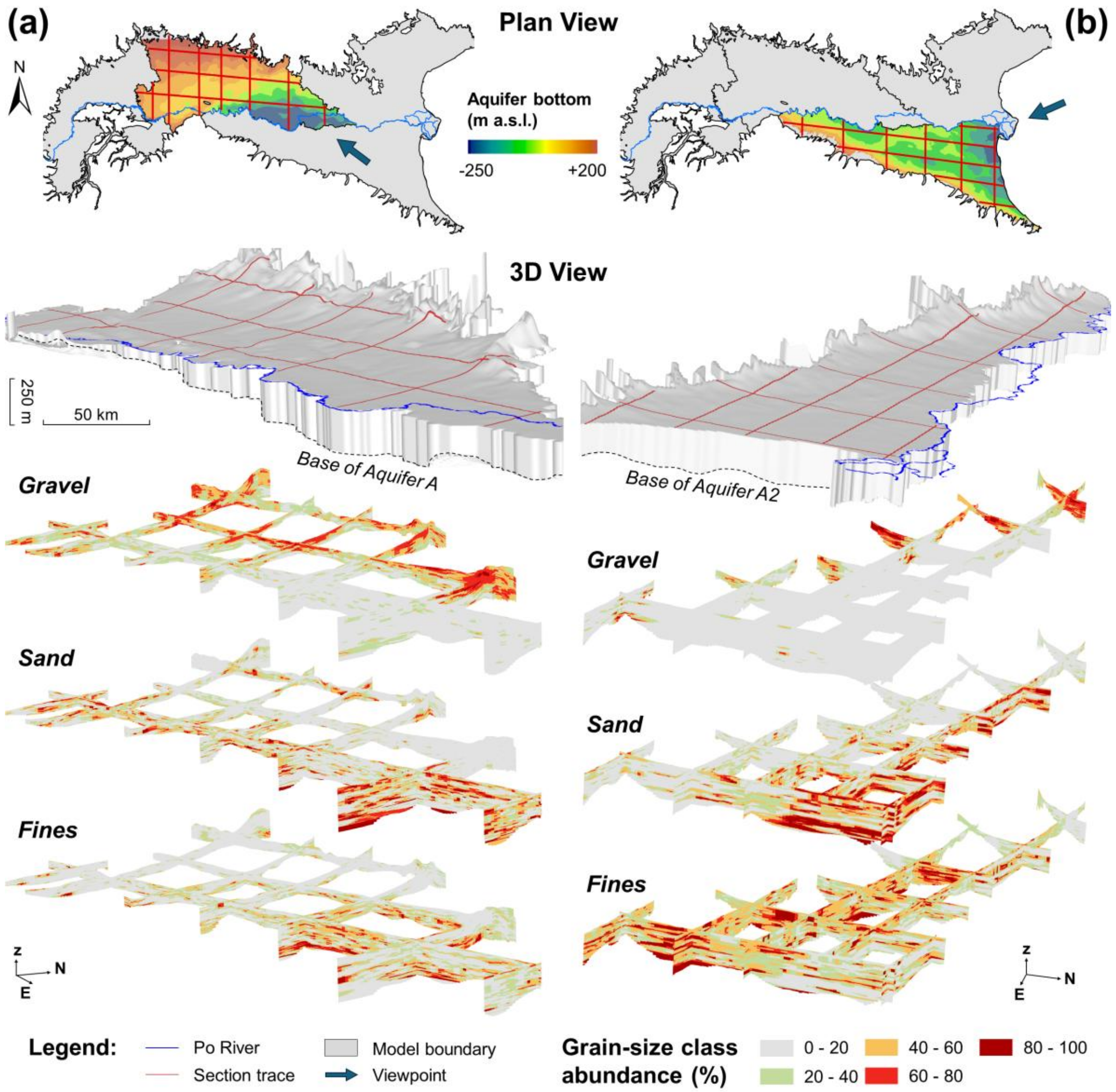
Methods

Results

Classification

Spatialization

Conclusions



3D hydrogeological model

3D interpolation (ordinary kriging) of grain size and permeability values attributed to borehole logs

- Constrained by hydrostratigraphic surfaces (aquifer boundaries)

Increased data density for modeling from 0.3 to 8.7 data/km^2

Leveraging Deep Learning and Natural Language Processing for hydrogeological insights from borehole logs

Previati et al. (2025) – Read the publication here: https://doi.org/10.1016/j.ejrh.2024.102157

Navigate -> 2 min FULL

Navigate ↓

Introduction

Study area

Methods

Results

Classification

Spatialization

Conclusions

Conclusions

- ✓ **High Accuracy**: The deep learning LSTM-based algorithm achieved a 97.4% classification accuracy for hydrofacies recognition from borehole log texts.
- ✓ **Enhanced Data Density**: Spatial data density improved significantly—from 0.34 to 8.7 data/km²—by integrating borehole descriptions with grain size and aquifer test data.
- ✓ **Consistent Validation**: Results showed strong agreement with existing hydrogeological maps, cross-sections, and literature data.
- ✓ **Algorithm Strengths**: Key success factors included a large training dataset, use of word embeddings, and detailed grain size references in structured logs.
- ✓ **Replicability**: The method is transferable to other regions with adequate borehole data and pre-labeled training sets.

Limitations

- Relies heavily on grain size descriptions, lacking a stratigraphic perspective.
- Cannot fully substitute for quantitative subsurface modeling or define hydrostratigraphic boundaries alone.
- The method is effective for mapping hydrofacies proportions within units but not for delineating stratigraphic boundaries.

Future Potential

Combining text-based classification with additional well logs could support more advanced hydrostratigraphic modeling.