Supplementary materials for "A physics-informed machine learning (PIML) framework for projecting 21st-century permafrost extent in Northeast China"

Shuai Huang[1, 2, 3], Xiangbing Kong[3 *], Xue Yang[1, 4, 5], Xiaoying Jin[1, 2], Shanzhen Li[4], Lin Yang[1, 4, 6], Yaodan Zhang[1, 4, 7], Kai Gao[8, 9], Hongwei Wang[8], Xiaoying Li[8,] Ruixia He[8], Lanzhi Lü[8], Guodong Cheng[8] and Huijun Jin[1, 2, 4, 5 *]

[1] *School of Ecology, Northeast Forestry University, Harbin 150080, China;*

[2] *Key Laboratory of Sustainable Forest Ecosystem Management (Ministry of Education), College of Forestry, Northeast Forestry University, Harbin 150080, China;*

[3] *Département de mathématiques, informatique et génie, Université du Québec à Rimouski, Rimouski G5L 3A1, Canada;*

[4] *School of Civil Engineering and Transportation, Permafrost Institute, and Ministry of Education Northeast-China Station of Permafrost Geo-environment, Northeast Forestry University, Harbin 150040, China;*

[5] *Ministry of Natural Resources Field Observation and Research Station of Permafrost and Cold Regions Environment in the Da Xing'anling Mountains at Mo'he, Northeast China, Natural Resources Survey Institute of Heilongjiang Province, Harbin 150036, China;*

[6] *College of Agriculture, Chifeng University, Chifeng 010019, China;*

[7] *College of Civil Engineering and Architecture, Quzhou Institute, Quzhou 324000, China*

[8] *State Key Laboratory of Cryospheric Science and Frozen Soil Engineering, Chinese Academy of Sciences, Lanzhou 730000, China*

[9] *University of Chinese Academy of Sciences, Beijing, 100049, China*

* *Corresponding authors: H Jin, School of Ecology, Northeast Forestry University, Harbin 150080, China, email: hjjin@nefu.edu.cn and; X Kong, Département de mathématiques, informatique et génie, Université du Québec à Rimouski, Rimouski G5L 3A1, Canada, email: xiangbing_kong@uqar.ca*

The supplementary materials provide detailed support for the study titled *"An optimized modeling scheme for projecting 21st-century permafrost extent in Northeast China."*

Figure S1 shows the spatial distribution of meteorological stations overlaid on a permafrost type base map from Ran et al. (2021).

Figure S2 illustrates the modeling framework of the physics-informed machine learning (PIML) approach used to simulate mean annual ground temperature (MAGT).

Figure S3 presents development potential maps for seven land use and land cover (LULC) types: cropland, forest, grassland, water body, barren land, impervious surface, and wetland.

Figure S4 shows the relationship between mean annual ground temperature (MAGT) and permafrost zonation index (PZI).

Table S1 quantifies the projected changes in atmospheric freezing index (AFI) and thawing index (ATI) under four SSP scenarios across 14 CMIP6 models, reporting mean, standard deviation, and coefficient of variation for each.

Table S2 lists the CMIP6 models used, including their modeling centers, land surface models, and atmospheric grid resolutions.

Table S3 enumerates natural, climatic, and socio-economic drivers of LULC change, with metadata on temporal coverage, spatial resolution, and data sources.

Table S4 evaluates the variable importance and prediction errors of each driver variable for different LULC classes based on the Land Expansion Analysis Strategy (LEAS) module of the Patch-generating Land Use Simulation (PLUS) model.

Table S5 provides class-specific parameter values for n_t, n_f, and r_k used in the TTOP model, referencing values from Zhang et al. (2022) and Obu et al. (2019).

Table S6 outlines the input features for the PIML model, including physiographic, soil, climatic, and surface variables, along with their spatial and temporal resolutions and sources.

Methods such as ArcGIS slope/aspect extraction and NDVI reconstruction *via* maximum value compositing are specified for data preprocessing. The materials conclude with full citations for referenced datasets and methodologies.

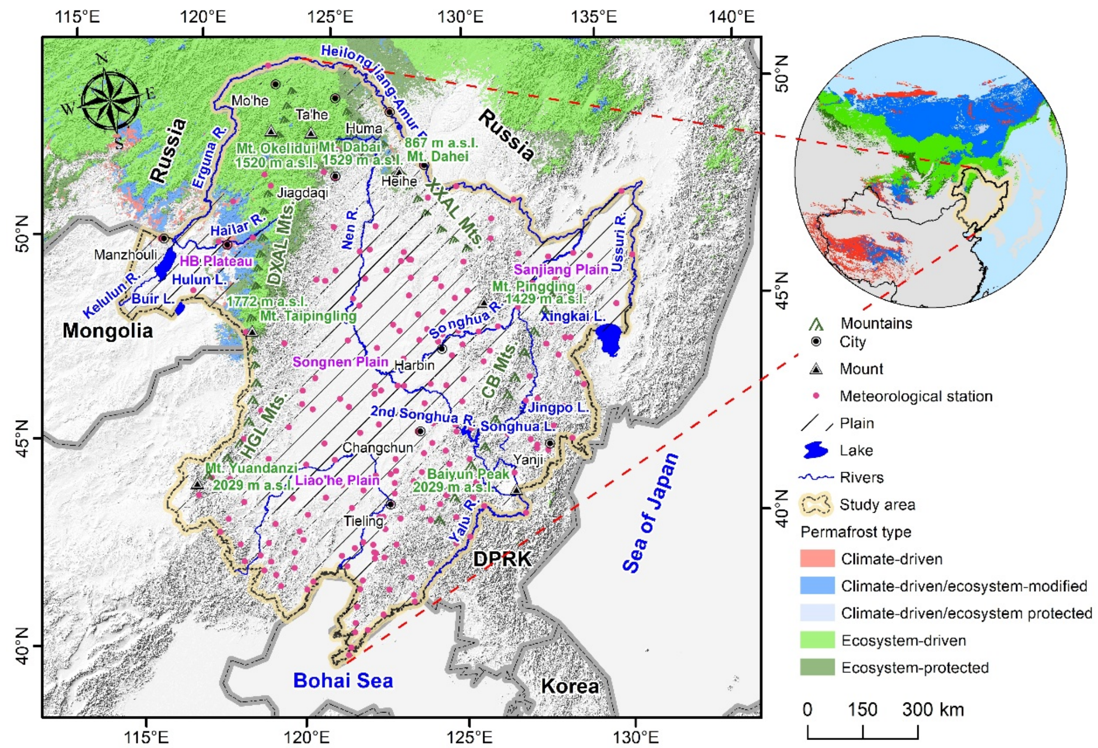

Figure S1. Location map of meteorological station in Northeast China (Base map of permafrost types derived and modified from Ran et al. (2021))

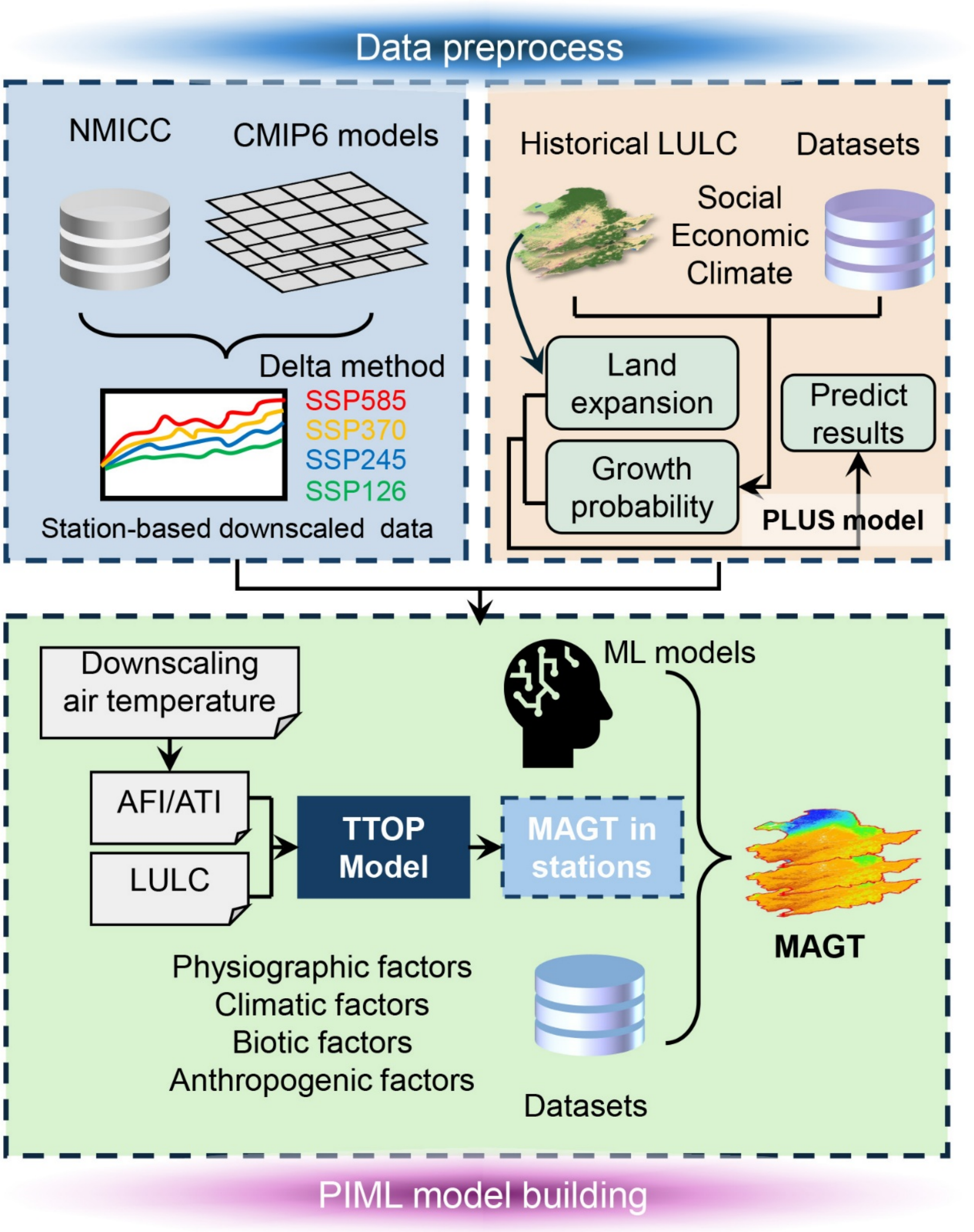

Figure S2. The framework of simulating MAGT based physics-informed machine learning (PIML) model and the modeling workflow.

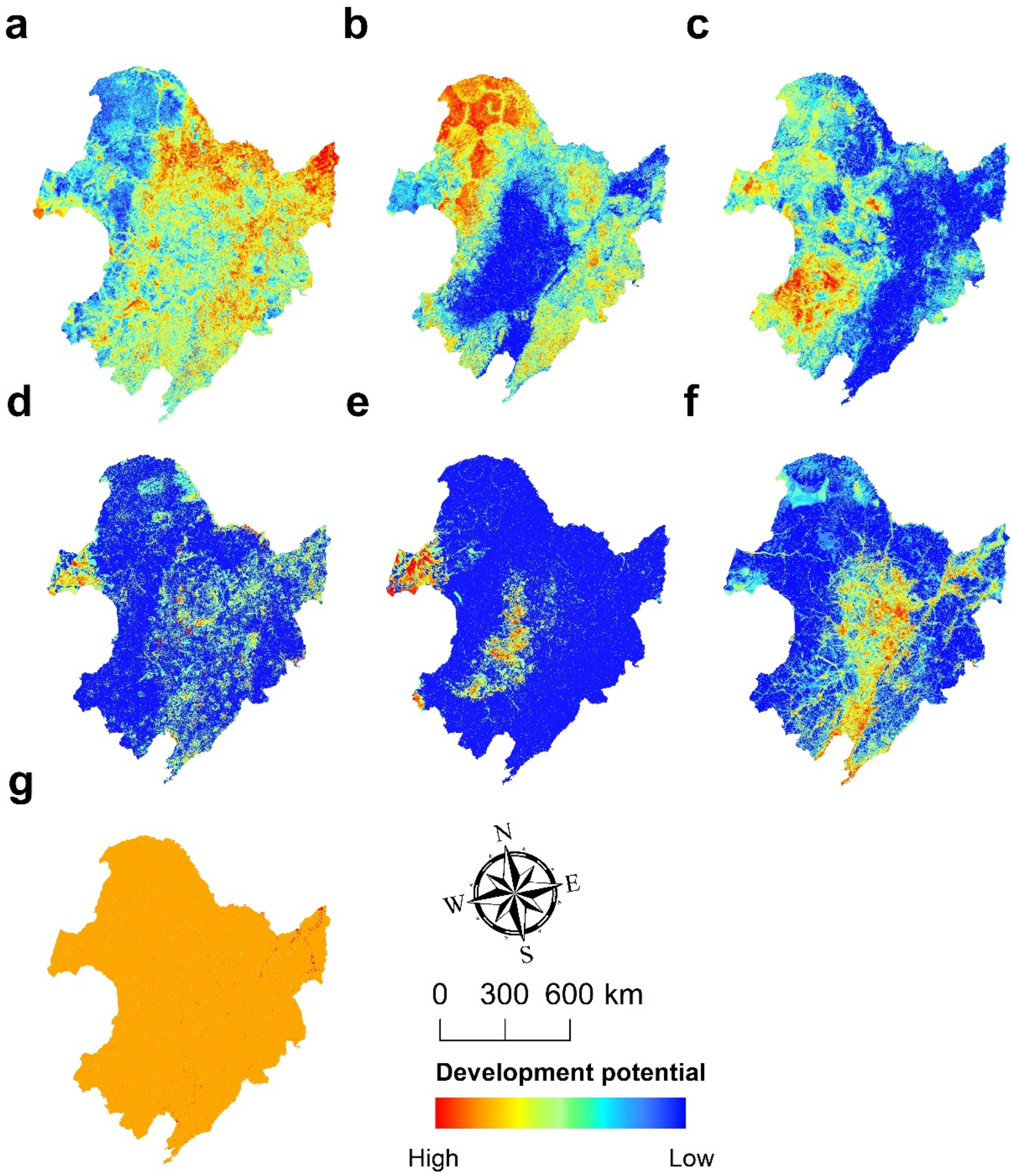

Figure S3. Development potential map for each land use and land cover type.

Notes: Panels (a–g) represents the development potential of (a) cropland, (b) forest, (c) grassland, (d) water body, (e) barren, (f) impervious surface, and; (g) wetland, respectively.

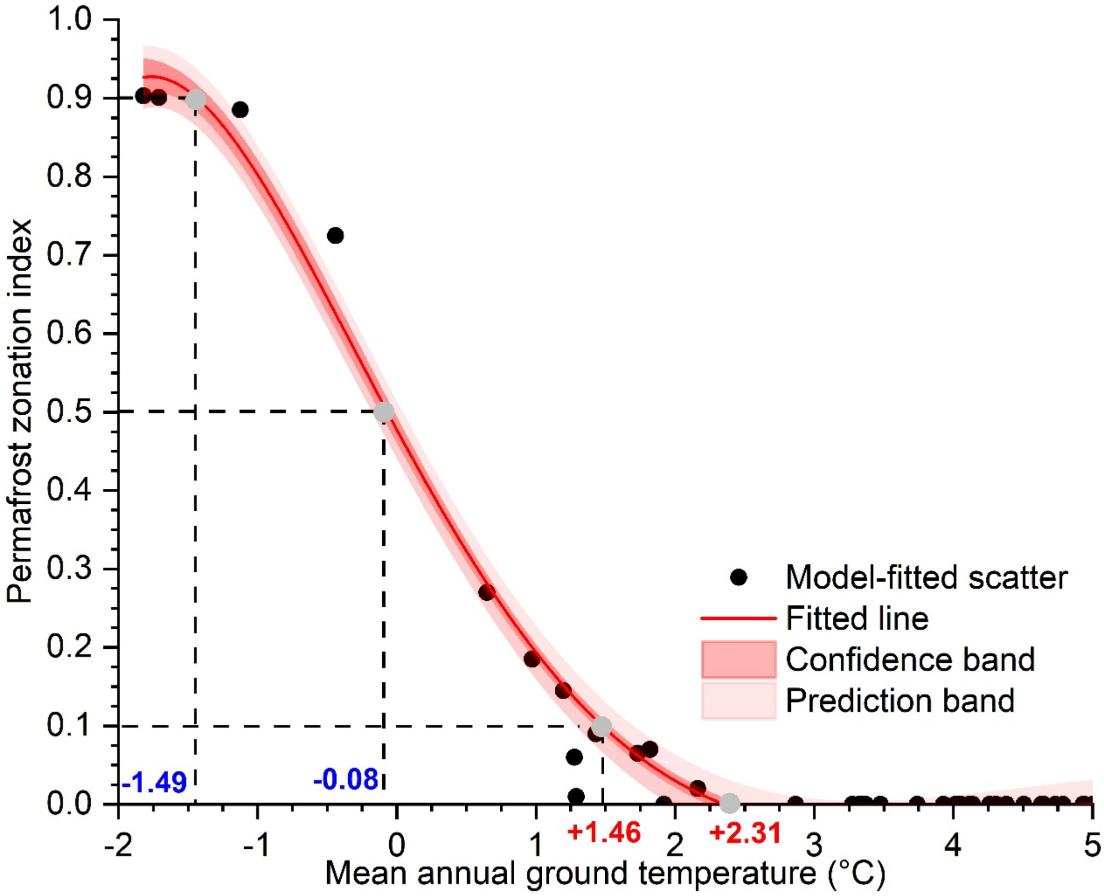

Figure S4. Relationship between mean annual ground temperature (MAGT) and permafrost zonation index (PZI)

Notes: A polynomial regression model was fitted to characterize the nonlinear relationship between the permafrost zonation index (PZI) and mean annual ground temperature (MAGT). Vertical dashed lines indicate the MAGT thresholds corresponding to PZI classifications: continuous ($PZI \geq 0.9$), discontinuous ($0.5 \leq PZI < 0.9$), sporadic ($0.1 \leq PZI < 0.5$), and isolated patches permafrost ($0.005 \leq PZI < 0.1$). The dark red band indicates the 95% confidence interval of the fitted curve, while the light red band represents the 95% prediction interval, reflecting the expected spread of individual PZI values around the fit.

Table S1. List of CMIP6 models used in this study

Models	Modeling Center	Land Surface Model	Grid Resolution (Atmosphere)
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Org. (CSIRO), Australia	CABLE2.0	1.875° × 1.25°
BCC-CSM2-MR	Beijing Climate Center (BCC), China	BCC-AVIM2.0	2.5° × 1.40625°
CanESM5	Canadian Centre for Climate Modelling and Analysis (CCCma), Canada	CLASS-CTEM3.6	~2.8° (T63 spectral)
CESM2-WACCM	National Center for Atmospheric Research (NCAR), USA	CLM5	0.9° × 1.25°
CMCC-ESM2	Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Italy	CLM4.5	0.9° × 1.25°
CNRM-CM6-1	Centre National de Recherches Météorologiques (CNRM) and CERFACS, France	SURFEX	~1.4° × 1.4°
CNRM-ESM2-1	CNRM and CERFACS, France	SURFEX	~1.4° × 1.4°
INM-CM4-8	Institute of Numerical Mathematics (INM), Russia	INM-Land	2.0° × 1.5°
INM-CM5-0	Institute of Numerical Mathematics (INM), Russia	INM-Land	2.0° × 1.5°
IPSL-CM6A-LR	Institut Pierre-Simon Laplace (IPSL), France	ORCHIDEE	2.5° × 1.25°
MIROC6	Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan	MATSIRO	1.4° × 1.4°
MRI-ESM2-0	Meteorological Research Institute (MRI), Japan	HAL	1.125° × 1.125°
NorESM2-LM	Norwegian Climate Centre (NCC), Norway	CLM5	1.9° × 2.5°
NorESM2-MM	Norwegian Climate Centre (NCC), Norway	CLM5	1.0° × 1.0°

Notes: ACCESS-ESM1-5: Australian Community Climate and Earth System Simulator – Earth System Model version 1.5; BCC-CSM2-MR: Beijing Climate Center Climate System Model version 2 – Medium Resolution; CESM2-WACCM: Community Earth System Model version 2 – Whole Atmosphere Community Climate Model; CMCC-ESM2: Centro Euro-Mediterraneo sui Cambiamenti Climatici Earth System Model version 2; CNRM-CM6-1: Centre National de Recherches Météorologiques Climate Model version 6.1; CNRM-ESM2-1: Centre National de Recherches Météorologiques Earth System Model version 2.1; CanESM5: Canadian Earth System Model; INM-CM4-8: Institute of Numerical Mathematics Climate Model version 4.8; INM-CM5-0: Institute of Numerical Mathematics Climate Model version 5.0; IPSL-CM6A-LR: Institute Pierre-Simon Laplace Climate Model version 6A – Low Resolution; MIROC6: Model for Interdisciplinary Research on Climate version 6; MRI-ESM2-0: Meteorological Research Institute Earth System Model version 2.0; NorESM2-LM: Norwegian Earth System Model version 2 – Low Resolution, and; NorESM2-MM: Norwegian Earth System Model version 2 – Medium Resolution.

Table S2. The driving factors of land use and land cover changes in this study.

Categories	Variables	Periods	Pixel resolution	Data resource
Natural factors	Elevation	-	30 m	https://lpdaac.usgs.gov
	Slope angle	-	30 m	
	Soil type	2000	1km	http://www.geodata.cn
	Types of soil erosion	2010	1km	https://www.resdc.cn
Climatic factors	Air temperature	2000-2020	1km	https://www.resdc.cn
	Precipitation			
Socio-economic	GDP	2020	1km	https://www.resdc.cn
	Population density			
Accessibility	Distance to expressway	2017	1km	https://www.ngcc.cn
	Distance to primary roads			
	Distance to secondary roads			
	Distance to tertiary roads			
	Distance to railway			
	Distance to water bodies			
	Distance to settlements			
	Distance to the government			

Notes: Slope angle was extracted using the slope analysis tool in ArcGIS 10.6, and both slope angle and elevation were resampled to a spatial resolution of 1 km. Distance metrics were calculated using the distance analysis tool in ArcGIS 10.6.

Table S3. The variable importance and prediction errors of driving factors for land use transitions based on the Land Expansion Analysis Strategy (LEAS) module of the Patch-generating Land Use Simulation (PLUS) model.

Factors	Cropland		Forest		Grassland		Water body		Barren		Impervious surface		Wetland	
	RMSE	IM	RMSE	IM	RMSE	IM	RMSE	IM	RMSE	IM	RMSE	IM	RMSE	IM
Distance to primary roads	0.27	0.06	0.23	0.06	0.19	0.05	0.11	0.07	0.07	0.05	0.18	0.06	0.01	0.00
Distance to secondary roads	0.27	0.06	0.21	0.05	0.20	0.06	0.08	0.02	0.08	0.08	0.16	0.04	0.01	0.04
Distance to tertiary roads	0.26	0.05	0.21	0.05	0.18	0.04	0.08	0.03	0.05	0.02	0.17	0.05	0.01	0.00
Distance to the government	0.25	0.05	0.22	0.05	0.19	0.05	0.10	0.05	0.06	0.04	0.20	0.08	0.01	0.10
Elevation	0.41	0.15	0.21	0.05	0.24	0.09	0.18	0.18	0.08	0.09	0.23	0.10	0.02	0.31
Types of soil erosion	0.20	0.01	0.16	0.02	0.16	0.02	0.06	0.01	0.08	0.09	0.13	0.01	0.01	0.00
GDP	0.27	0.06	0.21	0.05	0.20	0.06	0.13	0.11	0.09	0.10	0.21	0.09	0.01	0.01
Distance to expressway	0.27	0.06	0.20	0.04	0.20	0.06	0.08	0.02	0.06	0.04	0.19	0.06	0.02	0.13
Population density	0.27	0.06	0.20	0.04	0.20	0.06	0.18	0.18	0.06	0.04	0.31	0.18	0.03	0.41
Precipitation	0.28	0.07	0.26	0.09	0.34	0.18	0.09	0.04	0.10	0.12	0.16	0.04	0.01	0.00
Distance to railway	0.24	0.04	0.19	0.03	0.20	0.06	0.07	0.02	0.05	0.02	0.17	0.05	0.01	0.00
Distance to settlements	0.29	0.08	0.21	0.05	0.20	0.06	0.11	0.08	0.07	0.05	0.18	0.06	0.01	0.01
Distance to water bodies	0.23	0.03	0.19	0.03	0.22	0.07	0.07	0.02	0.08	0.08	0.14	0.02	0.01	0.01
Slope angle	0.29	0.07	0.41	0.19	0.19	0.05	0.11	0.08	0.10	0.13	0.21	0.08	0.01	0.00
Soil type	0.22	0.03	0.22	0.06	0.18	0.04	0.08	0.04	0.05	0.02	0.13	0.02	0.01	0.00
Air temperature	0.36	0.12	0.33	0.14	0.22	0.07	0.10	0.06	0.06	0.03	0.18	0.06	0.01	0.00

Notes: RMSE represents root mean square error, and; IM represents variable importance.

Table S4. List of n_t, n_f and r_k assigned to land use and land cover classes

LULC classes	n_t	n_f	r_k
Barren	1.26	0.44	0.95
Grassland	1.33	0.63	0.75
Cropland	1.30	0.43	0.75
Forest	1.21	0.39	0.95
Wetland	1.25	0.40	0.55
Impervious surface	1.26	0.44	0.70

Notes: n_f and n_t derived from Zhang et al. (2022); r_k derived from Obu et al. (2019)

Table S5. The driving factors of physics-informed machine learning (PIML) models in this study.

Categories	Variables	Temporal resolution	Pixel resolution	Data resource
Physiographic factors	Elevation Slope angle Slope aspect	-	30 m	https://lpdaac.usgs.gov
	Latitude Longitude		-	-
Soil properties	Bulk density Clay contents Silt contents Sand contents Organic carbon contents	-	250 m	https://soil.geodata.cn
Climatic factors	Land surface temperature	Daily	1 km	https://data.tpdc.ac.cn
	Snow cover duration	Yearly	1 km	
Surface characteristics	NDVI	10-days	1 km	https://www.resdc.cn
	LULC	10-years	30 m	https://zenodo.org

Notes: Slope angle and aspect were extracted using the slope and aspect analysis tools in ArcGIS 10.6. To mitigate the effects of atmospheric noise, cloud contamination, and variations in the solar zenith angle, a maximum value composite method was applied to reconstruct the NDVI datasets.

Table S6. The change trend of atmospheric freezing or thawing index (AFI/ATI) under four Shared Socioeconomic Pathways (SSPs).

AFI	SSP126	SSP245	SSP370	SP585
ACCESS-ESM1-5	-1.82	-3.80	-7.92	-7.85
BCC-CSM2-MR	-2.64	-5.18	-10.62	-9.48
CESM2-WACCM	-1.71	-4.42	-6.71	-9.04
CMCC-ESM2	-7.01	-8.49	-8.72	-14.04
CNRM-CM6-1	-1.88	-3.63	-8.12	-12.50
CNRM-ESM2-1	-4.67	-6.22	-9.03	-9.88
CanESM5	-3.28	-7.39	-12.07	-14.58
INM-CM4-8	0.12	-2.36	-5.51	-7.40
INM-CM5-0	-1.47	-4.55	-5.96	-6.67
IPSL-CM6A-LR	-3.01	-7.96	-11.36	-14.93
MIROC6	-1.60	-3.26	-6.24	-9.32
MRI-ESM2-0	0.96	-2.56	-5.25	-7.09
NorESM2-LM	-1.05	-3.43	-6.40	-8.45
NorESM2-MM	-1.01	-3.42	-7.89	-9.05
Mean	-2.15	-4.76	-7.99	-10.02
Standard deviation	1.97	2.00	2.18	2.83
Coefficient of variation	0.92	0.42	0.27	0.28
ATI	**SSP126**	**SSP245**	**SSP370**	**SP585**
ACCESS-ESM1-5	3.62	7.81	13.52	16.23
BCC-CSM2-MR	2.10	7.74	15.30	15.38
CESM2-WACCM	2.19	6.56	12.08	19.43
CMCC-ESM2	10.38	13.76	14.79	22.21
CNRM-CM6-1	2.56	7.74	14.90	19.7
CNRM-ESM2-1	3.92	8.53	13.71	16.67
CanESM5	4.07	10.64	20.69	27.46
INM-CM4-8	1.98	3.74	8.57	9.82
INM-CM5-0	1.69	4.88	7.24	9.83
IPSL-CM6A-LR	4.62	10.77	18.23	26.04
MIROC6	2.95	6.26	11.04	15.70
MRI-ESM2-0	0.22	4.47	7.99	11.45
NorESM2-LM	2.52	6.49	11.12	17.30
NorESM2-MM	2.96	7.50	12.88	15.90
Mean	3.27	7.64	13.00	17.37
Standard deviation	2.33	2.69	3.78	5.33
Coefficient of variation	0.71	0.35	0.29	0.31

References

Obu, J., Westermann, S., Bartsch, A., *et al*., 2019. Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km^2 scale. *Earth-Sci. Rev*. **193**, 299-316. DOI: 10.1016/j.earscirev.2019.04.023.

Ran, Y., Jorgenson, M. T., Li, X., *et al*., 2021. Biophysical permafrost map indicates ecosystem processes dominate permafrost stability in the Northern Hemisphere. *Environ. Res. Lett*. **16**(9), 095010. DOI: 10.1088/1748-9326/ac20f3.

Zhan, D., Man, H., Zang, S., *et al*., 2022. Simulation of frozen ground distribution in northeast China based on a surface frost number model. *J. Geogr. Sci*. **32**(8), 1581-1600. DOI: 10.1007/s11442-022-2011-8.