



**Article info**

**Type of article:**

Original research paper

**DOI:**

<https://doi.org/10.58845/jstt.utt.2026.vn.6.4.11-25>

**\*Corresponding author:**

Email address:

[tuannt94@utt.edu.vn](mailto:tuannt94@utt.edu.vn)

**Received:** 01/04/2026

**Received in Revised Form:**

21/04/2026

**Accepted:** 22/04/2026

## Multilayer Perceptron neural network modeling in developing land subsidence susceptibility map: A case study of Hanoi area, Vietnam

Nguyen Thanh Tuan<sup>1\*</sup>, Bui Thi Nhu<sup>1</sup>, Tran Thanh Hai<sup>2</sup>

<sup>1</sup>University of Transport Technology, 54 Trieu Khuc Street, Thanh Liet Ward, Hanoi, Vietnam

<sup>2</sup>Civil Aviation Authority of Vietnam, Hanoi, Vietnam

**Abstract:** Land subsidence is a geotechnical hazard that affects infrastructure and construction works, particularly in the context of rapid urbanization. Therefore, predicting and zoning areas susceptible to land subsidence has become an important research direction. In this study, a Multilayer Perceptron (MLP) model was employed to construct a land subsidence susceptibility map for the Hanoi area. Land subsidence data were identified using the PS-InSAR technique, combined with 14 conditioning factors selected as input variables, including groundwater, rainfall, distance to rivers and streams, distance to roads, distance to faults, land use/land cover (LULC), topographic wetness index (TWI), normalized difference vegetation index (NDVI), stream power index (SPI), geology, slope aspect, surface curvature, slope angle, and elevation. The results indicate that the model achieved good predictive performance, with AUC values of 0,957 for the training set and 0,944 for the testing set. In addition, groundwater and geology were identified as the two factors exerting the strongest influence on land subsidence susceptibility in the study area.

**Keywords:** Land subsidence; Multilayer neural network (MLP); Subsidence risk zoning map; GIS; Frequency ratio (FR).



Thông tin bài viết  
Dạng bài viết:  
Bài báo nghiên cứu

DOI:  
<https://doi.org/10.58845/jstt.utt.2026.vn.6.4.11-25>

\*Tác giả liên hệ:  
Địa chỉ Email:  
[tuannt94@utt.edu.vn](mailto:tuannt94@utt.edu.vn)

Ngày nộp bài: 01/04/2026  
Ngày nộp bài sửa: 21/04/2026  
Ngày chấp nhận: 22/04/2026

## Ứng dụng mô hình mạng nơ-ron đa lớp (MLP) trong xây dựng bản đồ phân vùng nguy cơ sụt lún khu vực Hà Nội, Việt Nam

Nguyễn Thanh Tuấn<sup>1\*</sup>, Bùi Thị Như<sup>1</sup>, Trần Thanh Hải<sup>2</sup>

<sup>1</sup>Trường Đại học Công nghệ GTVT, 54 Triều Khúc, Thanh Liệt, Hà Nội, Việt Nam

<sup>2</sup>Cục Hàng Không Việt Nam, Hà Nội, Việt Nam

**Tóm tắt:** Sụt lún là một tai biến địa kỹ thuật có ảnh hưởng đến cơ sở hạ tầng và các công trình xây dựng, đặc biệt trong bối cảnh đô thị hóa. Do đó, việc dự báo và phân vùng các khu vực có nguy cơ sụt lún trở thành một hướng nghiên cứu được quan tâm. Trong nghiên cứu này, mô hình Mạng nơ-ron nhiều lớp (MLP) được sử dụng để xây dựng bản đồ phân vùng nguy cơ sụt lún cho khu vực Hà Nội. Dữ liệu sụt lún được xác định bằng kỹ thuật PS-InSAR, kết hợp với 14 yếu tố ảnh hưởng được lựa chọn làm biến đầu vào, gồm: nước ngầm, lượng mưa, khoảng cách đến sông suối, khoảng cách đến đường giao thông, khoảng cách đến đứt gãy, hiện trạng sử dụng đất (LULC), chỉ số độ ẩm địa hình (TWI), chỉ số thực vật chuẩn hóa (NDVI), chỉ số sức mạnh dòng chảy (SPI), địa chất, hướng bờ dốc, hình dáng bề mặt, góc bờ dốc và độ cao địa hình. Kết quả cho thấy mô hình đạt hiệu quả dự báo tốt, với giá trị AUC đạt 0,957 trên dữ liệu huấn luyện và 0,944 trên dữ liệu kiểm chứng. Đồng thời, nước ngầm và địa chất được xác định là hai yếu tố ảnh hưởng lớn đến nguy cơ sụt lún tại khu vực nghiên cứu.

**Từ khóa:** Sụt lún đất; Hà Nội; Mạng nơ-ron nhiều lớp (MLP); Bản đồ phân vùng nguy cơ sụt lún; Hệ thống thông tin địa lý (GIS); Tỷ số tần suất (FR).

### 1. Mở đầu

Sụt lún là hiện tượng hạ thấp cao độ mặt đất theo phương thẳng đứng do tác động của các yếu tố tự nhiên và con người như sự nén chặt của các lớp đất trầm tích, khai thác nước ngầm, tải trọng công trình, khai thác tài nguyên hoặc sụp đổ các khoảng trống karst, cùng với sự biến đổi của điều kiện địa chất – địa chất thủy văn theo thời gian [1, 2]. Hiện tượng này ảnh hưởng đến hạ tầng kỹ thuật, công trình xây dựng và quá trình phát triển đô thị. Tại các đô thị lớn, đặc biệt ở khu vực đồng bằng và ven biển, sụt lún thường diễn ra trong thời gian dài làm gia tăng nguy cơ hư hỏng công trình

và chi phí duy tu, sửa chữa. Vì vậy, việc phân tích quy luật phân bố, diễn biến và các yếu tố liên quan đến sụt lún có ý nghĩa đối với công tác cảnh báo, quy hoạch hạ tầng và quản lý tài nguyên nước dưới đất.

Để theo dõi và đánh giá sụt lún, nhiều phương pháp đã được áp dụng, từ các kỹ thuật truyền thống như thủy chuẩn, GPS/GNSS và quan trắc mốc lún đến các công nghệ viễn thám radar như DInSAR, PS-InSAR và SBAS. Các phương pháp đo đạc truyền thống thường cho độ chính xác tốt nhưng chủ yếu phản ánh biến dạng tại các điểm rời rạc, chi phí triển khai lớn và khó bao quát trên

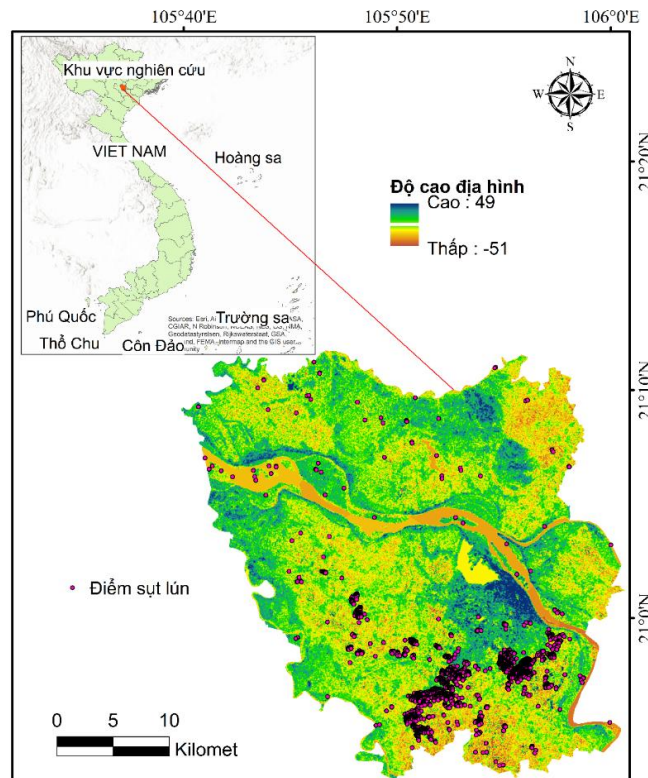
diện rộng [3, 4]. Trong khi đó, công nghệ viễn thám radar khẩu độ tổng hợp cho phép giám sát biến dạng bề mặt trên phạm vi rộng, trong nhiều điều kiện thời tiết và với độ phân giải không gian tương đối cao [5-7]. Trong các phương pháp, phương pháp giao thoa tán xạ ổn định (PS-InSAR) được sử dụng nhiều trong nghiên cứu đô thị do khai thác chuỗi ảnh SAR đa thời gian để xác định các điểm tán xạ bền vững và ước tính chuyển dịch bề mặt sau khi giảm ảnh hưởng của khí quyển và sai số quỹ đạo. Nhờ đó, phương pháp này phù hợp cho phân tích lún liên quan đến khai thác nước ngầm và các dạng biến dạng bề mặt khác.

Cùng với sự phát triển của công nghệ viễn thám, việc tích hợp dữ liệu biến dạng bề mặt với GIS và các thuật toán học máy, học sâu đã hình thành một hướng tiếp cận mới trong xây dựng bản đồ nguy cơ sụt lún [8]. Thay vì dựa trên các mô hình vật lý đòi hỏi tham số địa cơ học chi tiết thường khó thu thập ở quy mô đô thị, các thuật

toán như hồi quy logistic, máy vector hỗ trợ [9] hay mạng nơ-ron nhân tạo [10]... có khả năng học trực tiếp mối quan hệ phi tuyến giữa sụt lún và tập hợp các yếu tố điều kiện bao gồm địa hình, thạch học, hạ thấp mực nước ngầm, sử dụng đất và chỉ số thủy văn. Ưu điểm của hướng tiếp cận này không chỉ nằm ở độ chính xác dự báo không gian mà còn ở khả năng định lượng ảnh hưởng của các yếu tố.

Trên cơ sở đó, nghiên cứu này sử dụng kỹ thuật PS-InSAR để xây dựng dữ liệu hiện trạng biến dạng bề mặt, đồng thời tích hợp GIS với mô hình mạng nơ-ron nhiều lớp (MLP) để thành lập bản đồ phân vùng nguy cơ sụt lún. Cách tiếp cận này nhằm nhận diện các khu vực có nguy cơ lún và đánh giá vai trò tương đối của các yếu tố điều kiện, qua đó hỗ trợ phân tích không gian và cung cấp cơ sở tham khảo cho quy hoạch và quản lý hạ tầng trong khu vực nghiên cứu.

**2. Khu vực nghiên cứu**



**Hình 1.** Vị trí khu vực nghiên cứu và hiện trạng sụt lún.

Khu vực nghiên cứu là thành phố Hà Nội, thủ đô của Việt Nam, nằm ở trung tâm đồng bằng châu thổ sông Hồng, trong khoảng từ 20°53' đến 21°23'

vĩ độ Bắc và từ 105°44' đến 106°02' kinh độ Đông. Hà Nội có diện tích tự nhiên khoảng 3.359 km<sup>2</sup> và tiếp giáp với nhiều tỉnh thuộc vùng trung du và

đồng bằng Bắc Bộ (Hình 1). Với vị trí địa lý này, Hà Nội giữ vai trò là trung tâm chính trị, kinh tế, văn hóa và giao thông quan trọng của cả nước. Đồng thời, quá trình đô thị hóa nhanh, mở rộng xây dựng và gia tăng nhu cầu khai thác tài nguyên trong những năm gần đây cũng làm cho khu vực này trở thành đối tượng phù hợp để nghiên cứu các hiện tượng biến dạng bề mặt đất. Về địa hình, Hà Nội có sự phân hóa giữa khu vực đồi núi thấp ở phía bắc và tây bắc, tiêu biểu là Ba Vì và Sóc Sơn, với vùng đồng bằng thấp, tương đối bằng phẳng, chiếm phần lớn diện tích ở trung tâm, phía đông và phía nam thành phố.

Về địa chất, Hà Nội nằm trong bồn trũng Kainozoi sông Hồng, với các trầm tích Đệ Tứ có nguồn gốc sông, biển và đầm lầy, với thành phần chủ yếu là sét, bột, cát và cuội sỏi [11]. Các trầm tích Holocene và Pleistocene phân bố rộng rãi trong khu vực đồng bằng, trong khi ở các khu vực rìa như Ba Vì và Sóc Sơn xuất lộ các thành tạo địa chất cổ hơn. Về thủy văn, Hà Nội có mạng lưới sông ngòi phát triển, trong đó sông Hồng là trục chính, cùng với các sông Đáy, Nhuệ và Đuống, đóng vai trò quan trọng trong quá trình vận chuyển và bồi tụ trầm tích. Sự kết hợp giữa đặc điểm địa hình, địa chất và thủy văn đã tạo nên điều kiện tự nhiên đặc trưng của khu vực, đồng thời là cơ sở quan trọng để xem xét và phân tích hiện tượng sụt lún trong nghiên cứu này.

### 3. Thu thập và phân tích dữ liệu

#### 3.1. Hiện trạng sụt lún khu vực nghiên cứu

Bộ dữ liệu kiểm kê sụt lún đất trong nghiên cứu này được thành lập từ dữ liệu ảnh radar vệ tinh thông qua quy trình xử lý giao thoa radar đa thời gian nhằm nhận diện các vị trí có dấu hiệu lún sụt trong khu vực nghiên cứu. Từ kết quả xử lý, các điểm lún sụt được trích xuất, tổng hợp và chuẩn hóa để xây dựng bộ dữ liệu hiện trạng sụt lún đất, với tổng cộng 4.698 điểm được xác định [12]. Bộ dữ liệu này không chỉ phản ánh đặc điểm phân bố không gian của hiện tượng lún sụt mà còn đóng vai trò là cơ sở đầu vào quan trọng cho quá trình xây dựng mô hình phân vùng nguy cơ. Thông

qua bộ dữ liệu hiện trạng, nghiên cứu có điều kiện phân tích mối liên hệ giữa sự xuất hiện của lún sụt với các yếu tố điều kiện, từ đó phục vụ cho quá trình huấn luyện, kiểm chứng và đánh giá hiệu quả dự báo của mô hình.

#### 3.2. Các yếu tố ảnh hưởng đến sụt lún

Trong nghiên cứu này, 14 yếu tố ảnh hưởng đã được thu thập và phân tích, bao gồm hướng bờ dốc, hình dáng bề mặt, góc bờ dốc, độ cao địa hình, địa chất, khoảng cách đến đứt gãy, nước ngầm, lớp phủ/sử dụng đất (LULC), NDVI, lượng mưa, khoảng cách đến sông suối, khoảng cách đến đường giao thông, SPI và TWI (Bảng 1). Hướng bờ dốc ảnh hưởng đến sụt lún thông qua sự khác biệt về điều kiện bức xạ, độ ẩm và phong hóa bề mặt; phân tích FR cho thấy lớp hướng Bắc có giá trị cao nhất ( $FR = 1.137$ ). Hình dáng bề mặt ảnh hưởng đến sụt lún thông qua khả năng hội tụ hoặc phân tán dòng chảy bề mặt; phân tích FR cho thấy lớp lồi ( $>0.05$ ) có giá trị cao nhất ( $FR = 1.035$ ). Góc bờ dốc ảnh hưởng đến sụt lún thông qua trạng thái ổn định hình học của sườn dốc; phân tích FR cho thấy lớp 28.835–74.272 có giá trị cao nhất ( $FR = 10.871$ ). Độ cao địa hình ảnh hưởng đến sụt lún thông qua sự khác biệt về địa mạo, điều kiện thoát nước và phân bố vật liệu địa chất; phân tích FR cho thấy lớp 679–786 m có giá trị cao nhất ( $FR = 1.337$ ). Địa chất ảnh hưởng đến sụt lún thông qua đặc tính cơ lý và khả năng biến dạng của nền đất đá; phân tích FR cho thấy lớp trầm tích sông, suối, lũ tích và hỗn hợp có giá trị cao nhất ( $FR = 16.405$ ). Khoảng cách đến đứt gãy ảnh hưởng đến sụt lún thông qua mức độ nứt nẻ, tính thấm và sự suy yếu cấu trúc đất đá; phân tích FR cho thấy lớp 0–100 m có giá trị cao nhất ( $FR = 9.225$ ). Nước ngầm ảnh hưởng đến sụt lún thông qua sự thay đổi áp lực lỗ rỗng và quá trình nén lún của đất; phân tích FR cho thấy lớp -4.614 đến -0.735 có giá trị cao nhất ( $FR = 4.669$ ). Lớp phủ và sử dụng đất (LULC) ảnh hưởng đến sụt lún thông qua mức độ tác động của hoạt động bề mặt và điều kiện sử dụng không gian; phân tích FR cho thấy đất xây dựng có giá trị cao nhất ( $FR = 1.615$ ). NDVI ảnh hưởng đến sụt lún

thông qua mức độ che phủ thực vật và điều kiện ẩm bề mặt; phân tích FR cho thấy lớp 0.154–0.204 có giá trị cao nhất (FR = 2.401). Lượng mưa ảnh hưởng đến sụt lún thông qua sự thay đổi độ ẩm đất và trạng thái bão hòa của vật liệu nền; phân tích FR cho thấy lớp 1726.590–1798.180 mm có giá trị cao nhất (FR = 1.494). Khoảng cách đến sông suối ảnh hưởng đến sụt lún thông qua điều kiện thủy văn, xói mòn và tích tụ ẩm; phân tích FR cho thấy lớp 500–1000 m có giá trị cao nhất (FR = 2.474).

Khoảng cách đến đường giao thông ảnh hưởng đến sụt lún thông qua tải trọng công trình và mức độ tập trung hoạt động nhân sinh; phân tích FR cho thấy lớp 500–1000 m có giá trị cao nhất (FR = 1.288). SPI ảnh hưởng đến sụt lún thông qua khả năng tập trung dòng chảy bề mặt; phân tích FR cho thấy lớp 0–1687.627 có giá trị cao nhất (FR = 1.000). TWI ảnh hưởng đến sụt lún thông qua khả năng tích tụ ẩm của địa hình; phân tích FR cho thấy lớp 7.785–9.655 có giá trị cao nhất (FR = 1.076).

**Bảng 1.** Tỷ số tần suất của các yếu tố ảnh hưởng

Các yếu tố	Số lớp	Các lớp	Số điểm ảnh của các lớp	Số điểm ảnh sụt lún	Phần trăm điểm ảnh các lớp	Phần trăm điểm ảnh sụt lún	Tỷ số tần suất (FR)
<b>Hướng bờ dốc</b>	1	Mặt bằng	1092886	<b>904</b>	19.859	<b>19.242</b>	<b>0.969</b>
	2	Bắc	167904	163	3.051	3.470	1.137
	3	Đông Bắc	621478	461	11.293	9.813	0.869
	4	Đông	530196	425	9.634	9.046	0.939
	5	Đông Nam	616344	593	11.200	12.622	1.127
	6	Nam	565453	509	10.275	10.834	1.054
	7	Tây Nam	636568	564	11.567	12.005	1.038
	8	Tây	529458	459	9.621	9.770	1.016
	9	Tây Bắc	742958	620	13.500	13.197	0.978
<b>Hình dáng bề mặt</b>	1	Lõm (<-0.05)	1776734	1540	32.265	32.780	1.016
	2	Mặt bằng (-0.05 - 0.05)	1935265	1574	35.144	33.504	0.953
	3	Lồi (>0.05)	1794703	1584	32.591	33.716	1.035
<b>Góc bờ dốc</b>	1	0 - 2.912	3314075	2958	60.220	62.963	1.046
	2	2.912 - 7.572	2038340	1618	37.039	34.440	0.930
	3	7.572 - 16.310	142912	106	2.597	2.256	0.869
	4	16.310 - 28.835	7487	12	0.136	0.255	1.877
	5	28.835 - 74.272	431	4	0.008	0.085	10.871
<b>Độ cao địa hình</b>	1	291 - 570	1106759	617	20.098	13.133	0.653
	2	570 - 679	1936356	1825	35.164	38.846	1.105
	3	679 - 786	1627662	1856	29.558	39.506	1.337
	4	786 - 894	772781	377	14.033	8.025	0.572
	5	894 - 1008	63144	23	1.147	0.490	0.427
<b>Địa chất</b>	1	Hệ tầng Vĩnh Phúc	500	0	0.009	0.000	0.000
	2	Trầm tích sông, suối, lũ tích và hỗn hợp	37143	520	0.675	11.069	16.405
	3	Hệ tầng Nà Khuất	4218236	2121	76.623	45.147	0.589
	4	Hệ tầng Viên Nam	49832	552	0.905	11.750	12.980
	5	Hệ tầng Yên Duyệt	1191620	1500	21.645	31.928	1.475
	6	Hệ tầng Thạch Khoán	7875	5	0.143	0.106	0.744

Bảng 1. (tiếp)

Các yếu tố	Số lớp	Các lớp	Số điểm ảnh của các lớp	Số điểm ảnh sục lún	Phần trăm điểm ảnh các lớp	Phần trăm điểm ảnh sục lún	Tỷ số tàn suất (FR)
Khoảng cách đến đứt gãy	1	0 - 100	39006	307	0.708	6.535	9.225
	2	100 - 200	72644	411	1.319	8.748	6.632
	3	200 - 300	74907	519	1.360	11.047	8.121
	4	300 - 400	191042	1339	3.469	28.501	8.215
	5	400 - 500	5129143	2122	93.143	45.168	0.485
Nước ngầm	1	-22.363 - -14.488	136384	4	2.477	0.085	0.034
	2	-14.488 - -9.433	333760	817	6.063	17.390	2.868
	3	-9.433 - -4.614	406080	753	7.376	16.028	2.173
	4	-4.614 - -0.735	700736	2792	12.729	59.430	4.669
	5	-0.735 - 1.380	966784	168	17.561	3.576	0.204
	6	1.380 - 2.673	704768	41	12.802	0.873	0.068
	7	2.673 - 4.553	1385472	77	25.167	1.639	0.065
	8	4.553 - 6.317	773888	45	14.057	0.958	0.068
	9	6.317 - 7.727	97344	1	1.768	0.021	0.012
LULC	1	Đất nông nghiệp	1544410	86	28.047	1.831	0.065
	2	Đất xây dựng	3307575	4557	60.066	96.999	1.615
	3	Sông, ao, hồ	592927	50	10.768	1.064	0.099
	4	Thảm phủ thực vật	61627	5	1.119	0.106	0.095
NDVI	1	0.004 - 0.154	425639	84	7.731	1.788	0.231
	2	0.154 - 0.204	2012597	4124	36.556	87.782	2.401
	3	0.204 - 0.246	1632916	359	29.660	7.642	0.258
	4	0.246 - 0.283	1115522	107	20.262	2.278	0.112
	5	0.283 - 0.322	318855	24	5.792	0.511	0.088
Lượng mưa	1	1726.590 - 1798.180	3612800	4611	65.683	98.148	1.494
	2	1798.180 - 1873.639	1461376	74	26.569	1.575	0.059
	3	1873.639 - 1974.252	426176	13	7.748	0.277	0.036
Khoảng cách đến sông suối	1	0 - 300	721014	450	13.093	9.579	0.732
	2	300 - 500	282899	420	5.137	8.940	1.740
	3	500 - 1000	666613	1407	12.105	29.949	2.474
	4	1000 - 3000	1877054	1609	34.086	34.249	1.005
	5	>3000	1959162	812	35.578	17.284	0.486
Khoảng cách đến đường giao thông	1	0 - 100	614632	621	11.161	13.218	1.184
	2	100 - 300	1055459	845	19.167	17.986	0.938
	3	300 - 500	848696	835	15.412	17.774	1.153
	4	500 - 1000	1411826	1551	25.638	33.014	1.288
	5	>1000	1576129	846	28.622	18.008	0.629
SPI	1	0 - 1687.627	5502809	4698	99.992	100.000	1.000
	2	687.627 - 7875.596	386	0	0.007	0.000	0.000
	3	7875.596 - 19688.991	40	0	0.001	0.000	0.000
	4	19688.991 - 48378.663	4	0	0.000	0.000	0.000
	5	48378.663 - 144010.906	6	0	0.000	0.000	0.000

**Bảng 1. (tiếp)**

Các yếu tố	Số lớp	Các lớp	Số điểm ảnh của các lớp	Số điểm ảnh sput lún	Phần trăm điểm ảnh các lớp	Phần trăm điểm ảnh sput lún	Tỷ số tần suất (FR)
TWI	1	1.385 - 5.987	1720113	1455	31.256	30.971	0.991
	2	5.987 - 7.785	2150897	1926	39.084	40.996	1.049
	3	7.785 - 9.655	984937	905	17.897	19.264	1.076
	4	9.655 - 11.524	396055	351	7.197	7.471	1.038
	5	11.524 - 19.722	251243	61	4.565	1.298	0.284

**4. Phương pháp nghiên cứu**

**4.1. Mạng nơ-ron nhiều lớp (MLP)**

Mạng nơ-ron nhiều lớp được ứng dụng rộng rãi trong các bài toán phân loại nhị phân, bao gồm ba thành phần chính: lớp đầu vào, lớp ẩn và lớp đầu ra, trong đó các nơ-ron giữa các lớp liên tiếp được kết nối đầy đủ thông qua các trọng số có thể học được [13]. Trong nghiên cứu này, lớp đầu vào nhận véc-tơ đặc trưng  $t = (t_1, t_2, \dots, t_{14}) \in \mathbb{R}^{14}$  gồm 14 yếu tố ảnh hưởng đến hiện tượng sput lún đất, với kiến trúc một lớp ẩn được xác định qua quá trình tìm kiếm siêu tham số. Tín hiệu đầu vào được biến đổi qua lớp ẩn với hàm kích hoạt tanh (1)

$$h = \tanh(\mathbf{W}^{(1)}\mathbf{t} + \mathbf{b}^{(1)}) \tag{1}$$

trong đó  $\mathbf{W}^{(1)} \in \mathbb{R}^{1000 \times 14}$  và  $\mathbf{b}^{(1)} \in \mathbb{R}^{1000}$  lần lượt là ma trận trọng số và véc-tơ độ lệch của lớp ẩn. Hàm tanh được ưu tiên do có miền giá trị đối xứng trong khoảng  $(-1, 1)$ , giúp phân phối gradient đồng đều hơn, tăng tốc độ hội tụ và hạn chế hiện tượng triệt tiêu gradient (vanishing gradient). Đầu ra của lớp ẩn sau đó được ánh xạ sang xác suất phân loại qua hàm sigmoid theo công thức (2)

$$\hat{\phi} = \sigma(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)}) = \frac{1}{1 + e^{-(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)})}} \tag{2}$$

trong đó  $\hat{\phi} \in (0, 1)$  là xác suất dự báo khu vực có nguy cơ sput lún. Nhãn phân loại cuối cùng được xác định theo ngưỡng quyết định trong (3)

$$\phi = \begin{cases} 1 & \text{nếu } \hat{\phi} \geq 0,5 \text{ (sput lún)} \\ 0 & \text{nếu } \hat{\phi} < 0,5 \text{ (không sput lún)} \end{cases} \tag{3}$$

Quá trình huấn luyện được thực hiện qua hai giai đoạn: lan truyền xuôi (forward propagation) để tạo giá trị dự báo, và lan truyền ngược

(backpropagation) để cập nhật trọng số theo hướng cực tiểu hóa hàm mất mát Binary Cross-Entropy (BCE) theo công thức (4)

$$L(\mathbf{W}) = -\frac{1}{N} \sum_{i=1}^N [\phi_i \log(\hat{\phi}_i) + (1 - \phi_i) \log(1 - \hat{\phi}_i)] \tag{4}$$

Để hạn chế overfitting, chính quy hóa L2 với hệ số  $\alpha = 0.0001$  được tích hợp vào hàm mất mát theo công thức (5):

$$L_{reg}(\mathbf{W}) = L(\mathbf{W}) + \alpha \|\mathbf{W}\|^2 \tag{5}$$

**4.2. Phương pháp tỷ số tần suất (FR)**

Tỷ lệ tần suất là phương pháp định lượng được ứng dụng rộng rãi trong phân tích mối tương quan không gian giữa hiện tượng nghiên cứu và các yếu tố điều kiện liên quan [14]. Phương pháp này dựa trên nguyên lý so sánh xác suất xảy ra và không xảy ra của hiện tượng trong từng lớp phân loại của mỗi yếu tố, từ đó định lượng mức độ ảnh hưởng của từng lớp đến khả năng xuất hiện hiện tượng nghiên cứu.

Trong nghiên cứu này, FR được áp dụng để đánh giá mối quan hệ giữa phân bố không gian của hiện tượng sput lún đất và các yếu tố môi trường liên quan. Giá trị FR của từng lớp được xác định theo công thức (6):

$$FR = \frac{N_{sl}/N_t}{A_{cl}/A_t} \tag{6}$$

Trong đó  $N_{sl}$  là số điểm ảnh sput lún trong lớp phân loại thứ  $l$ ;  $N_t$  là tổng số điểm ảnh sput lún toàn khu vực;  $A_{cl}$  là số điểm ảnh của lớp phân loại thứ  $l$ ; và  $A_t$  là tổng số điểm ảnh toàn khu vực nghiên cứu.

**4.3. Phương pháp xác nhận**

Để đánh giá hiệu suất của mô hình, nghiên cứu sử dụng diện tích dưới đường cong (AUC) là

chỉ số nhằm định lượng khả năng phân biệt của mô hình giữa hai lớp sụt lún và không sụt lún [9]. Chỉ số này phản ánh mức độ mà mô hình có thể nhận dạng đúng các mẫu thuộc hai lớp trên toàn bộ ngưỡng phân loại nhị phân. Giá trị AUC biến thiên trong khoảng từ 0 đến 1, trong đó giá trị càng cao cho thấy mô hình có hiệu suất phân loại càng tốt và khả năng nhận dạng giữa hai lớp càng rõ ràng; ngược lại, giá trị càng thấp cho thấy năng lực phân biệt của mô hình càng hạn chế [15].

Giá trị AUC được xác định theo công thức (7):

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (7)$$

trong đó TPR tỷ lệ dương tính thực; FPR tỷ lệ dương tính giả.

Bên cạnh AUC, hiệu suất mô hình được đánh giá thông qua các chỉ số thống kê gồm độ nhạy (SST) (8), độ đặc hiệu (SPF) (9), giá trị dự đoán dương (PPV) (10), giá trị dự đoán âm (NPV) (11), độ chính xác (ACC) (12), hệ số Kappa (K) (13), sai số bình phương trung bình gốc (RMSE) (14) và sai số tuyệt đối trung bình (MAE) (15) [9]. Hệ số Kappa đánh giá mức độ đồng thuận giữa kết quả phân loại và nhãn thực tế, trong khi RMSE và MAE đo lường mức độ sai lệch trung bình giữa giá trị dự báo và thực tế, với giá trị càng nhỏ thì mô hình càng chính xác.

$$SST = \frac{TP}{TP+FN} \quad (8)$$

$$SPF = \frac{TN}{TN+FP} \quad (9)$$

$$PPV = \frac{TP}{TP+FP} \quad (10)$$

$$NPV = \frac{TN}{TN+FN} \quad (11)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$K = \frac{P_o - P_e}{1 - P_e} \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\phi_i - \hat{\phi}_i| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\phi_i - \hat{\phi}_i)^2} \quad (15)$$

trong đó N là tổng số mẫu,  $\phi_i$  là nhãn thực tế và  $\hat{\phi}_i$  là giá trị dự đoán của mẫu thứ i; TP, TN, FP, FN lần lượt là số điểm ảnh dương tính thật, âm tính

thật, dương tính giả và âm tính giả;  $P_o$  và  $P_e$  lần lượt là tỷ lệ đồng thuận thực tế và kỳ vọng ngẫu nhiên.

## 5. Kết quả và thảo luận

### 5.1. Hiệu suất mô hình

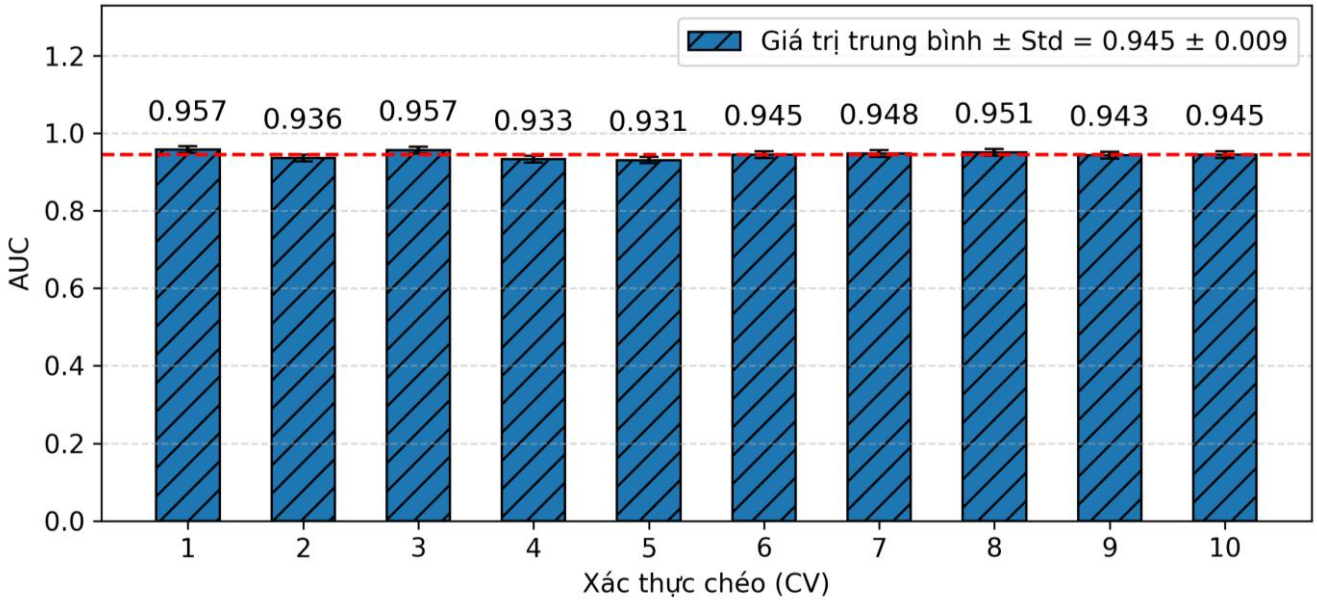
Dữ liệu trong nghiên cứu được phân chia theo tỷ lệ 70/30, trong đó 70% tổng số mẫu được sử dụng để xây dựng và huấn luyện mô hình, còn 30% số mẫu còn lại được sử dụng cho quá trình kiểm chứng mô hình nhằm đánh giá khả năng dự báo của mô hình trên tập dữ liệu chưa tham gia huấn luyện [16]. Trên dữ liệu huấn luyện, quá trình tối ưu hóa siêu tham số được thực hiện bằng phương pháp tìm kiếm lưới kết hợp xác nhận chéo 10 lần (GridSearchCV, k = 10) [17]. Cách tiếp cận này cho phép khảo sát có hệ thống nhiều tổ hợp siêu tham số khác nhau, từ đó lựa chọn cấu hình phù hợp nhất cho mô hình MLP.

Kết quả tối ưu cho thấy mô hình đạt hiệu quả tốt nhất với hàm kích hoạt tanh, hệ số chính quy hóa  $\alpha = 0.0001$ , kiến trúc một lớp ẩn gồm 1000 nơ-ron, cơ chế cập nhật tốc độ học constant và thuật toán tối ưu Adam. Kết quả đánh giá hiệu suất mô hình MLP qua 10 lần xác thực chéo được trình bày trong Hình 2. Giá trị AUC dao động trong khoảng từ 0.931 đến 0.957 với giá trị trung bình đạt  $0.945 \pm 0.009$ , cho thấy mô hình đạt hiệu suất ổn định và nhất quán qua tất cả các lần xác thực. Toàn bộ 10 giá trị AUC đều vượt ngưỡng 0,90 phản ánh khả năng phân biệt giữa khu vực sụt lún và không sụt lún ở mức độ rất cao. Độ lệch chuẩn thấp (Std = 0.009) cho thấy mô hình không bị phụ thuộc vào cách phân chia dữ liệu, từ đó khẳng định tính ổn định và khả năng tổng quát hóa tốt của mô hình trên tập dữ liệu nghiên cứu.

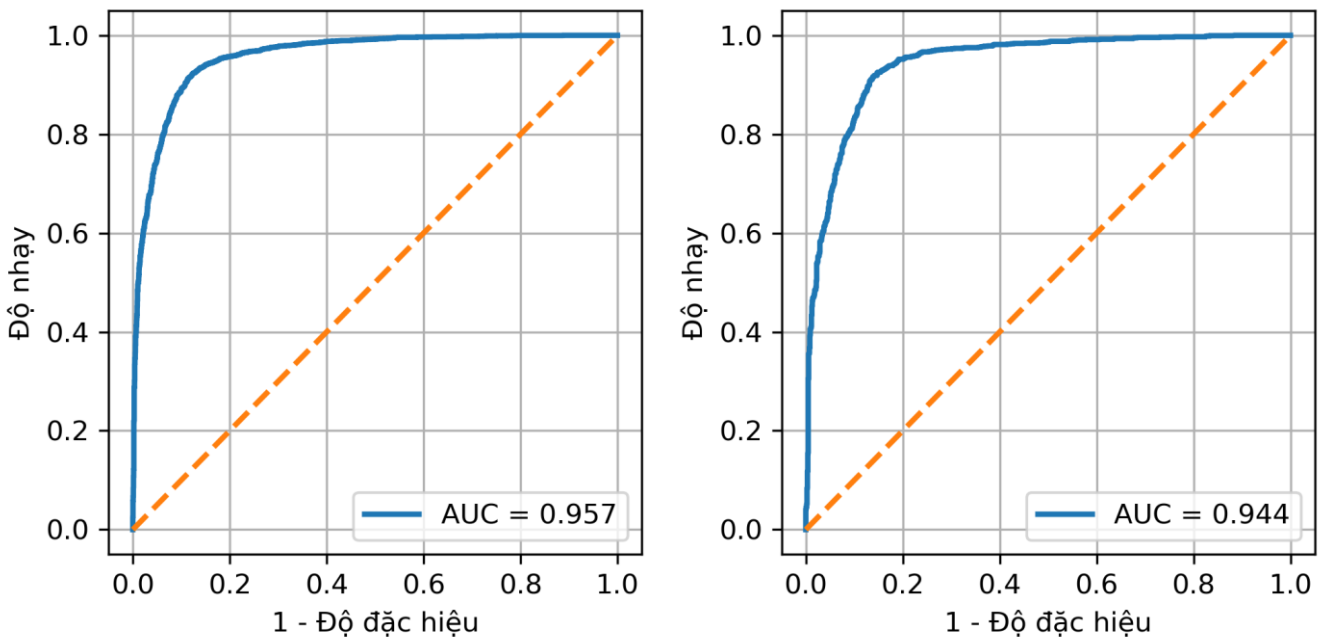
Kết quả đường cong ROC của mô hình MLP trên dữ liệu huấn luyện và dữ liệu kiểm chứng được trình bày trong Hình 3. Trên dữ liệu huấn luyện, mô hình đạt AUC = 0.957, trong khi trên dữ liệu kiểm chứng đạt AUC = 0.944, cho thấy khả năng phân biệt giữa khu vực sụt lún và không sụt lún tốt trên cả hai tập dữ liệu. Sự chênh lệch nhỏ giữa AUC dữ liệu huấn luyện và dữ liệu kiểm

chứng ( $\Delta = 0.013$ ) cho thấy mô hình không xảy ra hiện tượng overfitting, đồng thời khẳng định khả

năng tổng quát hóa tốt khi áp dụng trên dữ liệu chưa được huấn luyện.



Hình 2. Xác thực chéo (CV=10)



Hình 3. Giá trị AUC của mô hình MLP: (a) dữ liệu huấn luyện, (b) dữ liệu kiểm chứng.

Về khả năng phân loại, mô hình đạt độ chính xác tổng thể (ACC) lần lượt là 0.900 trên tập dữ liệu huấn luyện và 0.890 trên tập dữ liệu kiểm chứng, phản ánh hiệu suất dự báo cao và tương đối ổn định giữa hai tập dữ liệu (Bảng 2). Đối với độ đặc hiệu, chỉ số SPF đạt 0.881 và 0.866, cho thấy phần lớn các vị trí không sụt lún được nhận diện đúng; trong khi đó, độ nhạy (SST) đạt 0.917

và 0.910, chứng tỏ mô hình có khả năng phát hiện tốt các vị trí sụt lún. Giá trị dự đoán dương (PPV = 0.900 và 0.888) cùng với giá trị dự đoán âm (NPV = 0.900 và 0.892) đều ở mức cao, phản ánh độ tin cậy tốt của kết quả dự báo đối với cả hai nhóm vị trí sụt lún và không sụt lún. Xét theo các chỉ số dựa trên sai số, hệ số Kappa đạt 0.799 và 0.778 trên hai tập dữ liệu, cho thấy mức độ đồng thuận tốt

giữa kết quả phân loại của mô hình và dữ liệu thực tế.

Đồng thời, giá trị MAE = 0.100 và 0.110, cùng với RMSE = 0.316 và 0.332, cho thấy mức độ sai lệch dự báo tương đối thấp. Đặc biệt, diện tích dưới đường cong ROC (AUC) đạt 0.957 trên dữ liệu huấn luyện và 0.944 trên dữ liệu kiểm chứng,

khẳng định khả năng phân biệt rất tốt của mô hình giữa các vị trí sụt lún và không sụt lún. Sự chênh lệch nhỏ giữa các chỉ số trên dữ liệu huấn luyện và dữ liệu kiểm chứng cho thấy mô hình không có dấu hiệu overfitting rõ rệt, đồng thời khẳng định khả năng tổng quát hóa tốt trên dữ liệu chưa được huấn luyện.

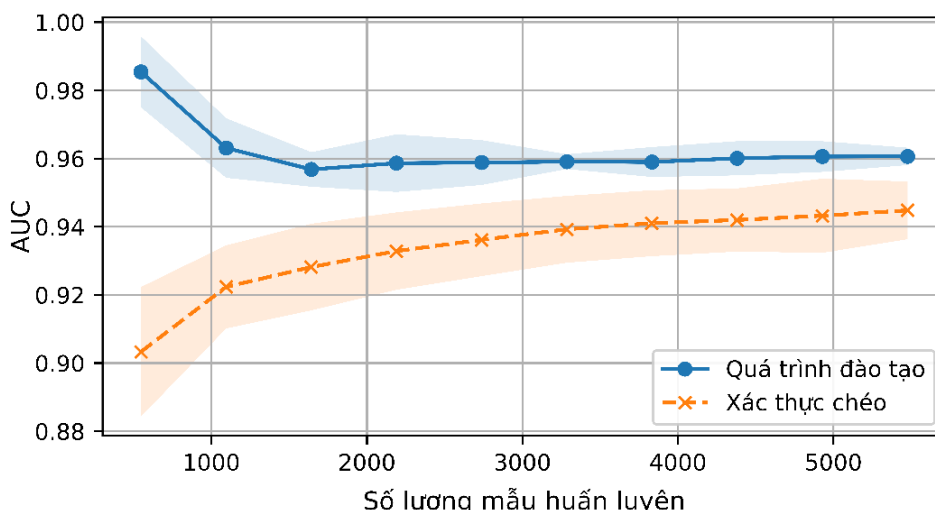
**Bảng 2.** Hiệu suất của mô hình

STT	Tham số	Dữ liệu huấn luyện	Dữ liệu kiểm chứng
1	TP	3019	1281
2	TN	2461	1041
3	FP	334	161
4	FN	272	126
5	PPV	0.9	0.888
6	NPV	0.9	0.892
7	SST	0.917	0.91
8	SPF	0.881	0.866
9	ACC	0.9	0.89
10	Kappa	0.799	0.778
11	MAE	0.1	0.11
12	RMSE	0.316	0.332
13	AUC	0.957	0.944

Ngoài ra, phân tích đường cong tích lũy (learning curve) của mô hình MLP nhằm đánh giá mối quan hệ giữa quy mô dữ liệu huấn luyện và hiệu suất mô hình (Hình 4).

Kết quả cho thấy khi số lượng mẫu huấn luyện tăng dần từ 500 đến 5500, AUC trên dữ liệu huấn luyện có xu hướng giảm nhẹ từ 0.986 xuống

và ổn định quanh mức 0.960, trong khi AUC trên tập xác thực chéo tăng dần từ 0.905 lên 0.945 và tiếp tục có xu hướng hội tụ. Khoảng cách giữa hai đường cong thu hẹp dần theo số lượng mẫu tăng lên, phản ánh mô hình đang dần khắc phục hiện tượng overfitting và cải thiện khả năng tổng quát hóa.



**Hình 4.** Đường cong tích lũy

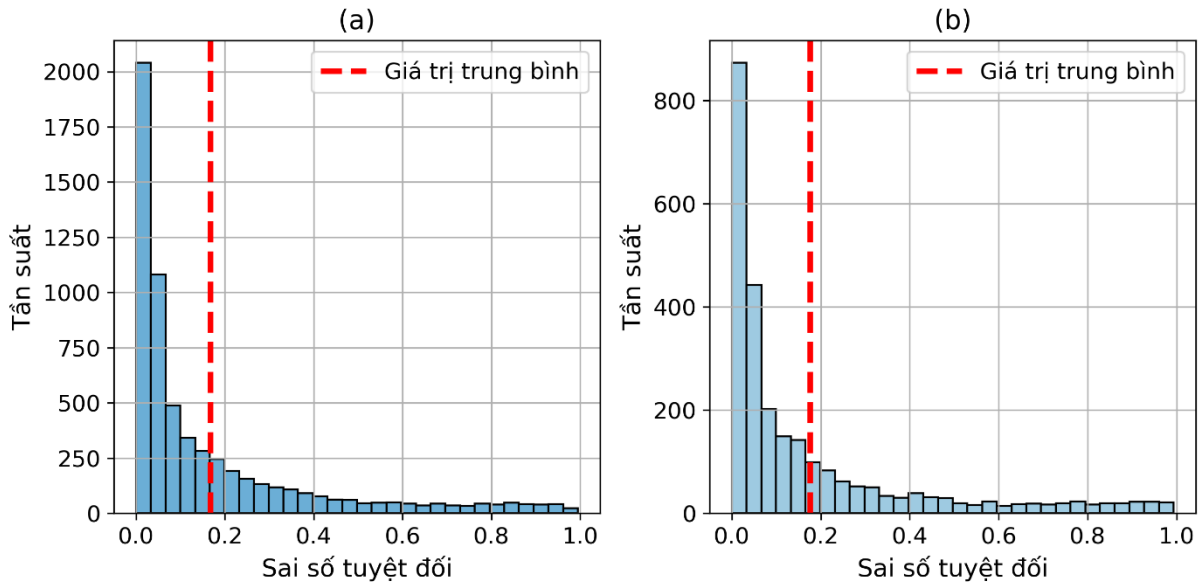
Phân bố sai số tuyệt đối của mô hình trên dữ liệu huấn luyện và dữ liệu kiểm chứng được trình bày trong Hình 5. Ở cả hai tập dữ liệu, phân bố sai

số đều tập trung ở vùng sai số thấp (0.0–0.2), cho thấy mô hình dự báo đúng hoặc gần đúng trên đa số trường hợp. Đáng chú ý, phân bố giữa dữ liệu

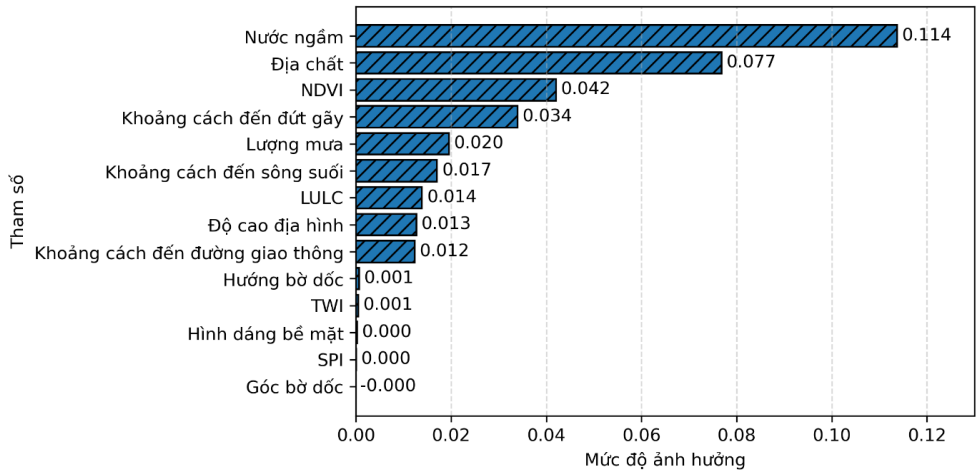
huấn luyện (a) và dữ liệu kiểm chứng (b) có sự tương đồng cao, với vùng sai số lớn có mức tần suất thấp, cho thấy các trường hợp dự báo sai lệch lớn chỉ chiếm tỷ lệ nhỏ và mô hình duy trì hiệu suất ổn định trên cả hai tập dữ liệu.

Mức độ ảnh hưởng của các yếu tố đầu vào được xác định thông qua phương pháp importance (Hình 6). Phương pháp này đo lường mức độ suy giảm hiệu suất mô hình, trong đó yếu tố nào làm AUC giảm nhiều hơn thì được xem là có tầm quan trọng cao hơn đối với kết quả dự báo [18]. Nhìn chung, các yếu tố có mức độ ảnh hưởng phân hóa rõ rệt, trong đó nước ngầm (0.114) và địa chất (0.077) là hai yếu tố có tác động lớn nhất, cho thấy điều kiện thủy địa chất đóng vai trò chi phối trong

quá trình hình thành sụt lún đất tại khu vực nghiên cứu. Tiếp theo, NDVI (0.042) và khoảng cách đến đứt gãy (0.034) thuộc nhóm có ảnh hưởng trung bình, phản ánh tác động của lớp phủ thực vật và cấu trúc địa chất kiến tạo đến sự ổn định của nền đất. Các yếu tố lượng mưa (0.020), khoảng cách đến sông suối (0.017), sử dụng đất (LULC = 0.014), độ cao địa hình (0.013) và khoảng cách đến đường giao thông (0.012) có mức ảnh hưởng thấp hơn nhưng vẫn đóng góp nhất định vào khả năng dự báo của mô hình. Đáng chú ý, các yếu tố hướng bờ dốc, TWI, hình dáng bề mặt, SPI và góc bờ dốc có mức độ ảnh hưởng gần bằng 0, cho thấy đóng góp của các yếu tố này vào kết quả phân loại là không đáng kể trong bối cảnh nghiên cứu.



Hình 5. Giá trị sai số: (a) dữ liệu huấn luyện; (b) dữ liệu kiểm chứng



Hình 6. Tầm quan trọng của biến

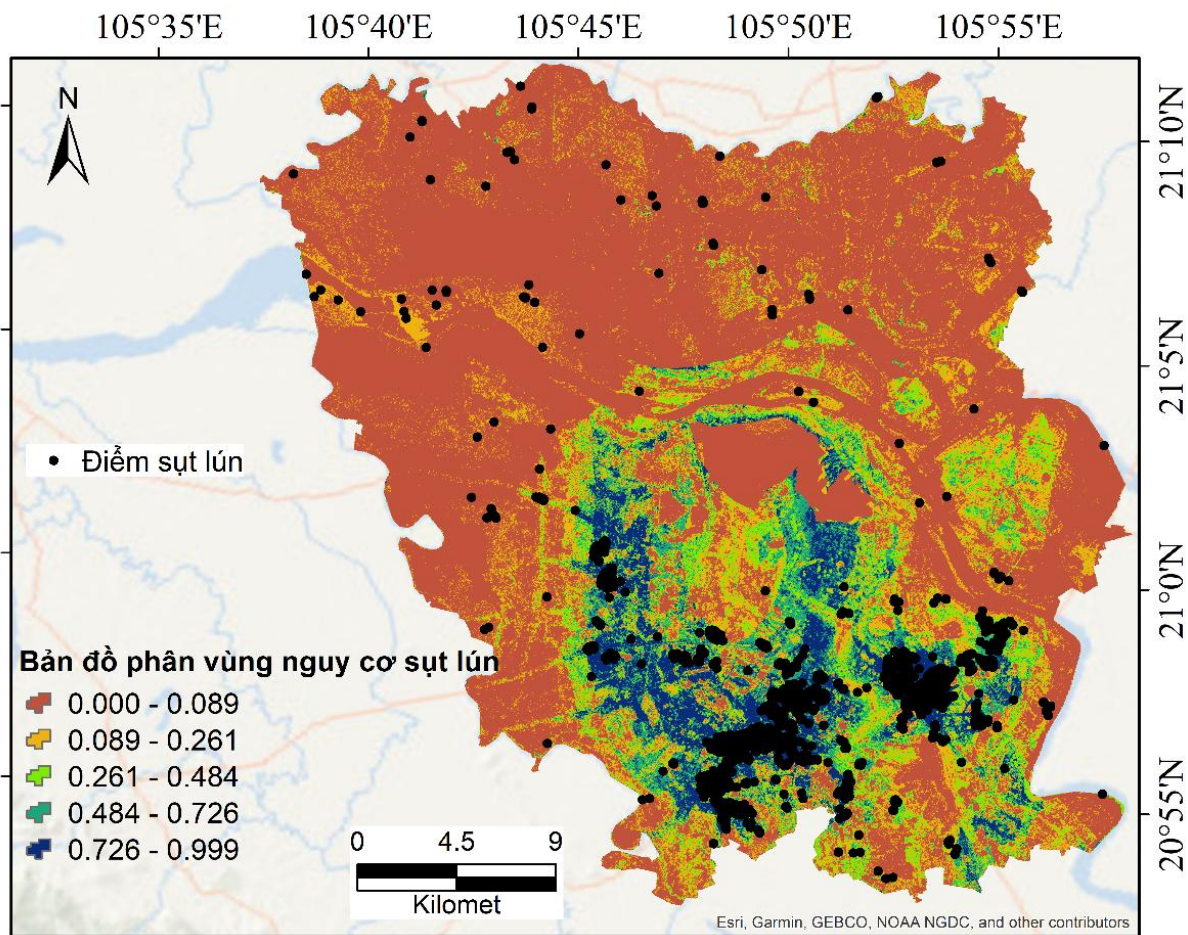
## 5.2. Xây dựng bản đồ phân vùng nguy cơ sụt lún khu vực nghiên cứu

### 5.2.1. Xây dựng bản đồ phân vùng nguy cơ sụt lún

Để thành lập bản đồ phân vùng nguy cơ sụt lún, giá trị xác suất dự báo từ mô hình tại từng vị trí không gian được xuất ra dưới dạng bảng dữ liệu, trong đó mỗi bản ghi lưu trữ tọa độ địa lý và giá trị xác suất sụt lún tương ứng [19]. Dữ liệu sau đó được tích hợp vào phần mềm ArcGIS 10.8 dưới dạng lớp điểm, phản ánh phân bố không gian của kết quả dự báo trên toàn khu vực nghiên cứu. Thông qua công cụ "Point to Raster", các giá trị xác suất rời rạc tại từng điểm được chuyển đổi thành lưới raster, tạo ra bề mặt không gian thể hiện sự biến thiên của nguy cơ sụt lún theo không gian địa lý, đồng thời tạo nền tảng cho quá trình phân cấp và trực quan hóa kết quả [20].

Quá trình phân cấp nguy cơ được thực hiện theo phương pháp Natural Breaks (Jenks) một phương pháp phân loại thống kê xác định ranh giới giữa các cấp dựa trên sự phân hóa tự nhiên trong phân bố dữ liệu [21]. Nguyên lý của phương pháp này là tối đa hóa sự khác biệt giữa các cấp trong khi giảm thiểu sự biến động nội bộ trong từng cấp, từ đó đảm bảo các ranh giới phân loại phản ánh trung thực cấu trúc thống kê của dữ liệu thay vì được áp đặt theo khoảng đều hoặc phân vị.

Kết quả phân loại cho ra năm cấp độ nguy cơ gồm rất thấp (0.000–0.089), thấp (0.089–0.261), trung bình (0.261–0.484), cao (0.484–0.726) và rất cao (0.726–0.999), tạo cơ sở khoa học cho việc phân tích và đánh giá phân bố không gian của nguy cơ sụt lún tại khu vực nghiên cứu một cách trực quan và có căn cứ thống kê (Hình 7).



Hình 7. Bản đồ phân vùng nguy cơ sụt lún khu vực nghiên cứu sử dụng mô hình MLP

### 5.2.2. Đánh giá độ tin cậy bản đồ phân vùng nguy cơ sụt lún

Để đánh giá mức độ phù hợp giữa kết quả phân vùng và phân bố thực tế của hiện tượng sụt lún, kết quả thống kê theo từng cấp nguy cơ được trình bày trong Bảng 3. Kết quả cho thấy sự phân hóa rõ rệt giữa các cấp, trong đó cấp rất cao (0.726–0.999) chiếm 78.46% số điểm sụt lún, với  $FR = 10.527$  cao hơn nhiều lần so với các cấp còn lại. Đối với phân cấp cao (0.484–0.726) chiếm

12.65% với  $FR = 1.953$ , tiếp tục thể hiện mối tương quan giữa cấp phân vùng và phân bố sụt lún. Cấp trung bình (0.261–0.484) chiếm 4.76% với  $FR = 0.597$ , phản ánh mức tương quan ở ngưỡng trung gian và đóng vai trò là vùng chuyển tiếp trong không gian phân vùng nguy cơ. Với hai cấp rất thấp và thấp mặc dù chiếm lần lượt 62.51% và 15.58% diện tích nghiên cứu nhưng mỗi cấp chiếm 2.06%, với  $FR$  lần lượt là 0.033 và 0.132, cho thấy xác suất xảy ra sụt lún tại hai vùng này là rất thấp.

**Bảng 3.** Phần trăm các vụ sụt lún và tỷ số tần suất của các lớp nhạy cảm sụt lún

Phân lớp	Tỷ số tần suất	Phần trăm điểm ảnh các lớp (%)	Phần trăm điểm ảnh sụt lún (%)
Rất thấp (0.000 - 0.089)	0.033	62.51	2.06
Thấp (0.089 - 0.261)	0.132	15.58	2.06
Trung bình (0.261 - 0.484)	0.597	7.98	4.76
Cao (0.484 - 0.726)	1.953	6.48	12.65
Rất cao (0.726 - 0.999)	10.527	7.45	78.46

Những kết quả trên cho thấy phù hợp với các nghiên cứu trước tại Hà Nội, cho thấy hiện tượng sụt lún không phân bố đồng đều mà thường tập trung rõ hơn ở khu vực phía nam, phía tây và một số vùng ven đô đang chịu tác động mạnh của quá trình đô thị hóa và khai thác nước ngầm. Các nghiên cứu InSAR trước đây tại Hà Nội như Đặng Vũ Khắc [12]... cũng đã ghi nhận xu hướng này, khi nhiều khu vực như Hà Đông, Hoài Đức và các vùng phát triển đô thị mở rộng thường xuất hiện lún rõ rệt hơn so với các khu vực ổn định khác. Do đó, việc phân lớn các điểm sụt lún trong nghiên cứu này tập trung trong các cấp nguy cơ cao và rất cao, đặc biệt cấp rất cao có  $FR = 10.527$ , cho thấy mô hình MLP phản ánh tốt quy luật phân bố không gian của hiện tượng sụt lún tại Hà Nội.

### 6. Kết luận

Nghiên cứu đã xây dựng bản đồ phân vùng nguy cơ sụt lún đất tại khu vực nghiên cứu thông qua mô hình mạng nơ-ron nhiều lớp (MLP) dựa trên 14 yếu tố điều kiện đầu vào và 4.698 điểm hiện

trạng sụt lún đã được thu thập. Kết quả cho thấy mô hình có khả năng dự báo tốt, với giá trị AUC đạt 0.957 trên dữ liệu huấn luyện và 0.944 trên dữ liệu kiểm chứng. Trong kết quả phân vùng, vùng nguy cơ rất cao chiếm 78.46% số điểm sụt lún, cho thấy bản đồ phân vùng có mức độ phù hợp tương đối tốt với hiện trạng sụt lún trong khu vực nghiên cứu. Đồng thời, kết quả nghiên cứu cũng cho thấy nước ngầm và địa chất là hai yếu tố ảnh hưởng lớn đến nguy cơ sụt lún, qua đó góp phần làm sáng tỏ cơ chế tác động của các điều kiện tự nhiên đối với sự hình thành và phân bố hiện tượng sụt lún đất trong khu vực nghiên cứu.

Tuy nhiên, nghiên cứu vẫn còn một số hạn chế nhất định, do bộ dữ liệu điểm sụt lún sử dụng trong nghiên cứu chưa phản ánh đầy đủ toàn bộ các sự kiện sụt lún đã xảy ra tại khu vực nghiên cứu, đồng thời nghiên cứu mới chỉ tập trung đánh giá hiệu quả của mô hình MLP mà chưa tiến hành so sánh với các mô hình học máy khác. Vì vậy, trong các nghiên cứu tiếp theo cần tiếp tục hoàn

thiện cơ sở dữ liệu hiện trạng sụt lún, bổ sung các yếu tố đầu vào và mở rộng đối sánh mô hình nhằm nâng cao độ chính xác, tính ổn định và giá trị ứng dụng của kết quả phân vùng.

### Lời cảm ơn

Tác giả xin gửi lời cảm ơn đến Trường Đại học Công nghệ Giao thông vận tải đã hỗ trợ tài chính cho nghiên cứu này trong khuôn khổ đề tài "Đánh giá nguy cơ sụt lún bề mặt đất sử dụng công nghệ trí tuệ nhân tạo kết hợp với các kỹ thuật địa không gian", mã số ĐTTĐ 2023-19.

### Tài liệu tham khảo

- [1] M. Younas et al. (2023). Geospatial analytics of driving mechanism of land subsidence in Gulf Coast of Texas, United States. *Science of the Total Environment*, 902, 166102. <https://doi.org/10.1016/j.scitotenv.2023.166102>
- [2] T. Nozadkhalil et al. (2023). Land subsidence due to natural gas extraction in the Thrace basin (NW Turkey) and its influence on the North Anatolian fault under the Marmara Sea. *Turkish Journal of Earth Sciences*, 32(SI-3). <https://doi.org/10.55730/1300-0985.1852>
- [3] H.Z. Abidin et al. (2001). Land subsidence of Jakarta (Indonesia) and its geodetic monitoring system. *Natural Hazards*, 365-387. <https://doi.org/10.1023/A:1011144602064>
- [4] H.Z. Abidin et al. (2010). Studying land subsidence in Semarang (Indonesia) using geodetic methods. *FIG Congress*
- [6] H. Sun et al. (2023). Land subsidence in a coastal city based on SBAS-InSAR monitoring: a case study of Zhuhai, China. *Remote Sensing*, 15(9), 2424. [https://doi.org/10.3390/rs15092424?urlappend=%3Futm\\_source%3Dresearchgate.net%26utm\\_medium%3Darticle](https://doi.org/10.3390/rs15092424?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle)
- [7] S.-H. Hong. (2024). Monitoring Time-Series Subsidence Observation in Incheon Using X-Band COSMO-SkyMed Synthetic Aperture Radar. *Korean Journal of Remote Sensing*, 40(2), 141-150. <https://doi.org/10.7780/kjrs.2024.40.2.2>
- [8] G. Wang et al. (2024). InSAR and machine learning reveal new understanding of coastal subsidence risk in the Yellow River Delta, China. *Science of the Total Environment*, 915, 170203. <https://doi.org/10.1016/j.scitotenv.2024.170203>
- [9] E. Hosseinzadeh et al. (2024). Evaluating machine Learning-Based approaches in land subsidence susceptibility mapping. *Land*, 13(3), 1-27. <https://doi.org/10.3390/land13030322>
- [10] A.A. Nadiri et al. (2021). Mapping risk to land subsidence: Developing a two-level modeling strategy by combining multi-criteria decision-making and artificial intelligence techniques. *Water*, 13(19), 2622. [https://doi.org/10.3390/w13192622?urlappend=%3Futm\\_source%3Dresearchgate.net%26utm\\_medium%3Darticle](https://doi.org/10.3390/w13192622?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle)
- [11] T. Nghi et al. (1991). Quaternary sedimentation of the principal deltas of Vietnam. *Journal of Southeast Asian Earth Sciences*, 6(2), 103-110. [https://doi.org/10.1016/0743-9547\(91\)90101-3](https://doi.org/10.1016/0743-9547(91)90101-3)
- [12] D.V. Khac et al. (2014). Recent land subsidence caused by the rapid urban development in the Hanoi region (Vietnam) using ALOS InSAR data. *Natural Hazards and Earth System Sciences*, 14(3), 657-674. <https://doi.org/10.5194/nhess-14-657-2014>
- [13] S. Piramuthu, M.J. Shaw, J.A. Gentry. (1994). A classification approach using multi-layered neural networks. *Decision Support Systems*, 11(5), 509-525. [https://doi.org/10.1016/0167-9236\(94\)90022-1](https://doi.org/10.1016/0167-9236(94)90022-1)
- [14] S.I. Elmahdy et al. (2020). Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain, UAE. *Geocarto International*, 37(1), 315-331. <https://doi.org/10.1080/10106049.2020.1716398>
- [15] J. Huang, C.X. Ling. (2005). Using AUC and accuracy in evaluating learning algorithms.

- IEEE Transactions on knowledge and Data Engineering*, 17(3), 299-310. <https://doi.org/10.1109/TKDE.2005.50>
- [16] I. Muraina. (2022). Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. *7th international Mardin Artuklu scientific research conference*.
- [17] A. Arabameri et al. (2022). Application of novel ensemble models and k-fold CV approaches for Land subsidence susceptibility modelling. *Stochastic Environmental Research and Risk Assessment*, 36(1), 201-223. <https://doi.org/10.1007/s00477-021-02036-7>
- [18] A.H. Sung. (1998). Ranking importance of input parameters of neural networks. *Expert systems with Applications*, 15(3-4), 405-411. [https://doi.org/10.1016/S0957-4174\(98\)00041-4](https://doi.org/10.1016/S0957-4174(98)00041-4)
- [19] G. Herrera-García et al. (2021). Mapping the global threat of land subsidence. *Science*, 371(6524), 34-36. <https://doi.org/10.1007/s00477-021-02036-7>
- [20] W. Bajjali. (2023). ArcGIS Pro and ArcGIS Online: Applications in Water and Environmental Sciences. Springer, 183-221.
- [21] C. Jia et al. (2021). Monitoring analysis and numerical simulation of the land subsidence in linear engineering areas. *KSCE Journal of Civil Engineering*, 25(7), 2674-2689. <https://doi.org/10.1007/s12205-021-1823-x>