

ARTIFICIAL NEURAL NETWORK FOR PREDICTION OF LAND SUBSIDENCE IN MUDSLIDES REGION THROUGH INSAR AND RAIN DATA

Anselmo Bettio^{1,2,*}, Francesco Sansone¹, Alessandro Francesconi^{1,2,3}

¹ Stellar Project srl

² Centro di Ateneo di Studi e Attività Spaziali “G. Colombo” (CISAS)

³ Department of Industrial Engineering, University of Padova (IT)

* Corresponding author

KEY WORDS: InSAR PSI, Data fusion, Time series analysis, LSTM, Mudslides

ABSTRACT:

Mudslides are powerful and fast-moving mass movements that pose significant risks to human lives, infrastructure, and natural environments. They are commonly triggered by intense rainfall and their impact is particularly severe in mountainous regions. Synthetic Aperture Radar (SAR) technology can be used to calculate the subsidence of the territory over time by means of a temporal series of SAR images through the Persistent Scatterer Interferometry (PSI) technique. In some research Interferometric SAR (InSAR PSI) data were used to train Long Short-Term Memory (LSTM) based Artificial Neural Network (ANN) to provide movements forecasting. This paper proposes a new LSTM based ANN to forecast future territory movements considering both the past InSAR PSI data, the rain forecasting of the next acquisition and the past cumulative amount of rain since the movements of mudslides are strictly dependent to the quantity of rainfall accumulated in the terrain. The results of the proposed ANN are shown in terms of Mean Square Error (MSE) and Mean Absolute error (MAE) by comparing them with a LSTM-based ANN trained with only the InSAR PSI data.

1. INTRODUCTION.

1.1. Motivation

Mudslides are powerful and fast-moving mass movements that pose significant risks to human lives, infrastructure, and natural environments. These hazardous events are commonly triggered by intense rainfall and can travel at high velocities, often exceeding 10 meters per second (Chen et al., 2017), making them one of the most dangerous types of mass movements. Improving the ability to forecast mudslide movements can help to minimize the risks associated with these destructive mass movements. An example of mudslide consequences is shown in Figure Figure 1.

Rainfall causes changes in surface and groundwater dynamics that reduce the slope stability conditions and cause mudslides. However, there is not an analytical relationship between the quantity of rainfall and the movement of the mudslide risk area. In fact, in this work, we are going to focus on the opposite of an analytical approach, those are Artificial Neural Networks, which are instead black boxes trained with a large dataset to find patterns

in the data, and to make predictions on future movements. In this work we want to

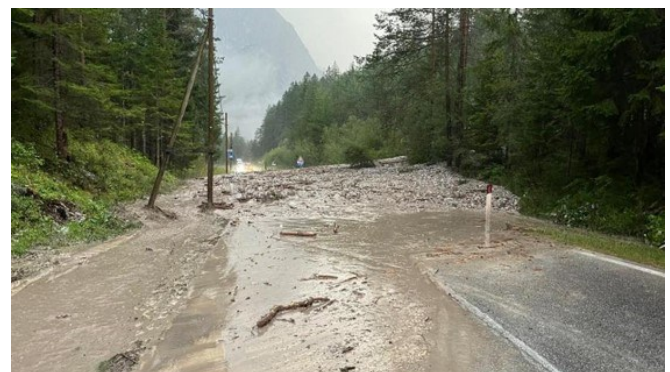


Figure 1. Example of mudslide in Acquabona locality.

1.2. InSAR PSI

InSAR Persistent Scatterer Interferometry (InSAR PSI) is an advanced technique used for monitoring and measuring ground deformation. This technique exploits the backscattered radar signals from persistent scatterers (PS), which are characterized by stable and long-lasting reflectivity properties. The InSAR PSI technique analyses a time series of SAR images acquired over an extended period, typically months to years. By tracking the phase evolution of the radar signals from these PS, InSAR PSI enables the precise measurement of surface deformation rates with millimeter-level accuracy (Ferretti et al., 2001).

1.3. State of the art

Some scientific works present the potentiality of statistical and machine learning methods to make prediction on future movements of the territory (Fiorentini et al., 2020) and (Naghbi et al., 2022). Some important studies have explored the possibility to forecast InSAR data by means of Long Short-Term Memory Recurrent Neural Networks (LSTM RNN) models (Chen et al., 2021), (Agrawal, 2022), (Hashemi et al., 2022) and (Hill et al., 2021). The LSTM RNN architecture has proven to be highly effective in capturing and modeling complex sequential patterns. It can process input sequences of varying lengths and can handle both short-term and long-term dependencies.

1.4. Objectives

The aim of this work is to evaluate and compare a LSTM RNN model for improving predictions of subsidence in mudslides risk areas with respect to the traditional LSTM RNN by integrating the large geological dependence of mudslides on heavy rainfall in the model. The LSTM RNN is designed in such a way to be trained considering a fusion of InSAR PSI data and meteorological data. The considered meteorological data are both the past cumulative amount of rain and the rain forecasting of the next acquisition. Then, it is compared with a traditional LSTM RNN trained with only InSAR PSI data. The remainder of the paper is structured as follows: Section 2 explains the followed methodology; Section 3 describes the chosen case study and the corresponding used data; in Section 4, the results in terms of predictors performance are presented; some concluding remarks and future works are discussed in Section 5.

2. METHODOLOGY

The methodology of this work is divided into the following steps: (1) design the two considered LSTM RNN models (i.e., with and without inclusion of rainfall data); (2) processing the Single Look Complex (SLC) SAR images to obtain the InSAR PSI data; (3) processing of the InSAR PSI data to obtain suitable training and test sets for the two LSTM RNN; (4) training, test and comparison of the two LSTM RNN.

2.1. LSTM RNN design

Two recurrent neural networks based on LSTM layers have been designed to make predictions on future movements of the territory subject to mudslides risks. The two networks differ from each other by the input that is provided to them. In particular, the first neural network (LSTM RNN 1) bases its predictions only on the past InSAR PSI data. In the second neural network (LSTM RNN 2), the strong dependence of the mudslide movements on the amount of rainfall is exploited to improve the forecasts. In fact, LSTM RNN 2 is designed to be trained with a data fusion of InSAR PSI displacement data and meteorological rain data taken in the same periods in which the InSAR measurements were made and the rain forecast of the future measurement to be done. Both LSTM RNN 1 and LSTM RNN 2 feature a parametric structure with two degrees of freedom: the number of LSTM layers and the number of units for each of these layers. A LSTM layer is an RNN layer that learns long-term dependencies between time steps in time series and sequence of data. The LSTM layer consists of LSTM cells equal to the size of the considered data time series window. Figure 2 shows the LSTM layer considered in this work, which contains 6 LSTM cells, since the data time series window comprises 6 consecutive measurements (i.e., the past five [InSAR data, Rain] pairs and the next acquisition rain forecasting).

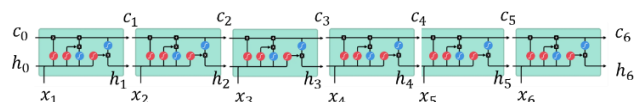


Figure 2. LSTM layer composed of six LSTM cells

Each of the LSTM cells are initialized with a cell state c_t and a hidden state $h_t, \forall t \in [0,6]$. Iteratively, the observation x_t , the previous hidden state h_{t-1} and the cell state c_t are combined through the operations shown in Figure 3.

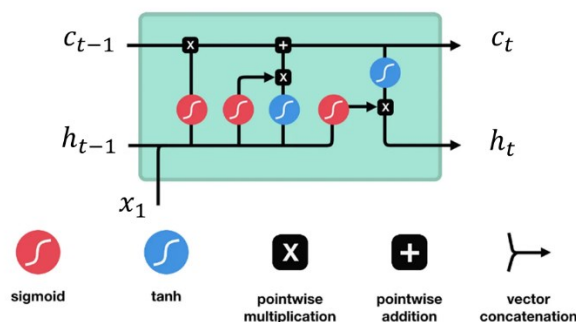


Figure 3. LSTM cell and operations in which x_t , h_{t-1} , c_t are involved

In TensorFlow, the parameter indicates the dimension of the hidden state vector is called “units”. By tuning the number of LSTM layers and the number of units, the optimization of the models, in terms of Mean Square Error (MSE, Section 2.4.1) and Mean Absolute Error (MAE, Section 2.4.2) is reached.

2.2. InSAR PSI processing

In this step, the InSAR PSI data is calculated by using the two software libraries called snap2stamps (Blasco and Foumelis, 2018) and StaMPS (Hooper et al., 2018). The snap2stamps library is composed of a series of Python scripts that use the SNAP ESA software API to perform processing on SAR images to make them ready for StaMPS processing, which calculates the movements data. Their combined use produces as output millimetric-accurate measurements on some points called Persistent Scatterers (PS), by receiving as input a time series of at least 15 SAR images.

2.3. InSAR PSI post-processing

In this step, the Persistent Scatterers generated in the previous one are processed through automated Python scripts that perform the following functions:

- For each PS:
- - A normalization of the displacement measurements with respect to the initial instant of time is performed.
 - A moving average over 5 consecutive measurements, which are acquired with a rate of 6 days. Thus, each measure is the mean movement of the last 30 days (6 days/measurement * 5 measurements = 30 days) is performed in such a way to reduce the uncertainty of the displacement measurements to ± 3 mm with respect to 8 mm of uncertainty on the single measurement (Marinkovic et al., 2007).
 - For the periods of time covered by the SAR acquisitions, the rain meteorological data (ARPAV, ARPAV website) are considered. A moving average of the amount of accumulated rain over 30 consecutive days is performed in such a way to have the same measurement rate calculated for the InSAR PSI data.
 - Finally, for each of the two considered LSTM RNN the training and test sets are prepared. The pair (features, label) used to form the dataset of the LSTM RNN 1 has the following structure:

$$([A_{-4}, A_{-3}, A_{-2}, A_{-1}, A_0], [A_1])$$

The pair (features, label) used to form the dataset of the LSTM RNN 2 add the rainfall information to the structure:

$$\left(\left[\begin{array}{c} [A_{-4}, R_{-4}], [A_{-3}, R_{-3}], [A_{-2}, R_{-2}], [A_{-1}, R_{-1}], \\ [A_0, R_0], [R_1] \end{array} \right], [A_1] \right)$$

Where, $A_i \forall i \in [-4, 1]$ is the InSAR PSI measurement at acquisition time i . $R_i, \forall i \in [-4, 1]$ is the rain measurement at acquisition time i . Note: acquisition time $i = 1$ represents the future time at which both the InSAR PSI data has to be predicted by the model and the rain forecasting is provided to the model.

2.4. Train and evaluation of LSTM RNN performances

The previously designed LSTM RNN 1 and 2 are designed with two degrees of freedom: the number of LSTM layers and the number of units for each of these layers. These parameters are tuned to find an optimum model in terms of MSE and MAE. The evaluation of each model is done with both the training set (i.e., 2000 samples) and a test set (i.e., 500 samples) containing new data never seen by the model.

In the following, the indexes used to evaluate our predictors (i.e., the designed neural networks) are described.

2.4.1. Mean Square Error (MSE)

In statistics, the mean squared error (MSE) of a predictor measures the average of the squares of the errors—that is, the average squared difference between the predicted values and the actual values (1).

$$MSE = \frac{1}{n} \sum_{k=0}^n (Y_k - \hat{Y}_k)^2 \quad (1)$$

Where, Y_k is the observed measurement and \hat{Y}_k is the predicted measurement.

2.4.2. Mean Absolute Error (MAE)

The MAE is defined as:

$$MAE = \frac{1}{n} \sum_{k=0}^n (Y_k - \hat{Y}_k) \quad (2)$$

Where, Y_k is the observed measurement and \hat{Y}_k is the predicted measurement.

3. CASE STUDY

3.1. Area of Interest

In this paper, a case study in Acquabona, in the heart of the Belluno Dolomites in northern Italy, is presented, which is considered a mudslides high risk area. In fact, several mudslides have already caused extensive damage in terms of infrastructure but above all human lives. Figure 4 shows the area taken into consideration in this research during the period from 30/04/2015 to 13/05/2017.



Figure 4. Area of interests subject to mudslides in in Acquabona (Belluno, Italy)



Figure 5. Persistent scatterers (PS) obtained after the snap2stamps + stamps workflow in Acquabona site. The PS time series within the red circled area are used in the proposed workflow

3.2. SAR data

The Single Look Complex (SLC) SAR data are the input of the first step described in Section 2.2, in which the InSAR PSI data are calculated. Considering the selected area of interest, the corresponding Sentinel-1 SLC SAR images in the period from 30/04/2015 to 13/05/2017 have been downloaded from the Copernicus ESA Open Access Hub website.

3.3. Meteorological data

The LSTM RNN 2 model considers both InSAR PSI displacement data and the meteorological rainfall data, which were downloaded from the Regional Agency for Environmental Prevention and Protection of the Veneto (ARPAV) website.

4. RESULTS

The data presented in Section 3 – Case Study are the input of the proposed workflow presented in Section 2 - Methodology. Therefore, the raw SAR data is transformed into InSAR PSI measurements with snap2stamps and StaMPS procedure obtaining the PS of the area of interest, which is circled in red in Figure 4.

After this, during the InSAR PSI post-processing step the data are prepared (i.e., normalization, noise removal, data structure handling) to be suitable for the train and test of the LSTM RNN 1 and LSTM RNN 2, which is performed.

As said, the two neural networks models are designed in parametric way, thus they can be trained and test by tuning two degrees of freedom parameters: the number of LSTM layers and units for each of these layers. For each LSTM RNN (i.e., 1 and 2) and for each combination of parameters (i.e., LSTM layer = {1, 2}; units per layer = {16, 32, 64}) training and test procedure of the neural networks are performed. To globally evaluate the several neural networks, the average performance of the predictions in terms of MSE and MAE in both training and test sets were calculated and reported in Table 1 and Table 2.

LSTM RNN 1 (trained with only InSAR data)	Evaluation with training set [MSE, MAE]	Evaluation with test set [MSE, MAE]
LSTM layers = 1 Units = 16	8.5, 2.4	7.2, 2.2
LSTM layers = 1 Units = 32	7.5, 2.2	5, 1.8
LSTM layers = 1 Units = 64	7.3, 2.2	7.7, 2.3
LSTM layers = 2 Units = 16	8.5, 2.4	6.5, 2
LSTM layers = 2 Units = 32	7.3, 2.2	8.1, 2.2
LSTM layers = 2 Units = 64	6.7, 2	7.6, 2.2

Table 1. Evaluation performances on training set and test set of the LSTM RNN 1 (trained with only InSAR data) in terms of mean square error (MSE) and mean absolute error (MAE) by varying the number of units and the number of LSTM layers.

LSTM RNN 2 (trained with InSAR + Rain data)	Evaluation with training set [MSE, MAE]	Evaluation with test set [MSE, MAE]
LSTM layers = 1 Units = 16	2.4, 0.9	9.7, 1.9
LSTM layers = 1 Units = 32	2.1, 0.8	9.7, 1.8
LSTM layers = 1 Units = 64	1.8, 0.7	9.9, 1.9
LSTM layers = 2 Units = 16	2.6, 0.9	6.8, 2.4
LSTM layers = 2 Units = 32	1.8, 0.7	8, 1.6
LSTM layers = 2 Units = 64	2, 0.7	4.4, 1.3

Table 2. Evaluation performances on training set and test set of the LSTM RNN 2 (trained with InSAR + Rain data) in terms of mean square error (MSE) and mean absolute error (MAE) by varying the number of units and the number of LSTM layers.

By looking at Table 1 and Table 2 some noticeable results of this work are listed:

- The best performances for the LSTM RNN 1 tested with new data are obtained when the network has one LSTM layer formed by 32 units. In this case the (MSE, MAE) pair is equal to (5 mm², 1.8 mm), which is within the uncertainty range of the InSAR PSI measure (9 mm², 3mm).
- The best performances for the LSTM RNN 2 tested with new data are obtained when the network has two LSTM layers each formed by 64 units. In this case the (MSE, MAE) pair is equal to (4.4 mm², 1.3 mm), which is within the uncertainty range of the InSAR PSI measure (9mm², 3mm).
- Considering the two proposed neural network with the best choice of parameters:
 - LSTM RNN 1 (layers=1, units=32)
 - LSTM RNN 2 (layers=2, units=64)

The proposed LSTM RNN 2 model trained with the data fusion of InSAR PSI and meteorological data improves the forecast performances of the 12% in terms of MSE and of the 27% in terms of MAE with respect to the traditional LSTM RNN 1 trained with only InSAR PSI data.

Figure 6 shows an example of predictions comparison between the two proposed models for a specific PS over the area of interest.

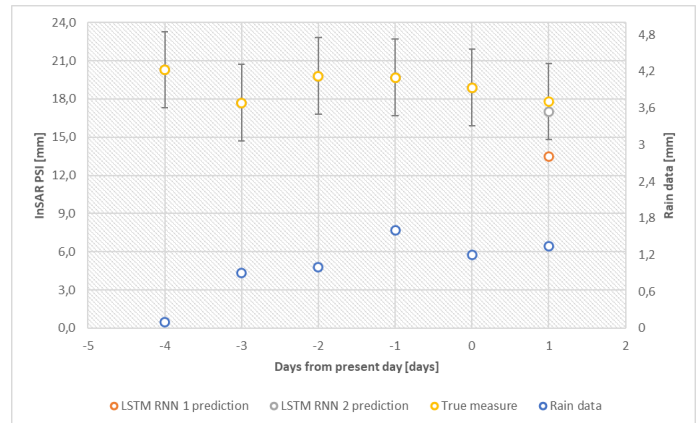


Figure 6. Predictions comparison between LSTM RNN 1 and LSTM RNN 2 models

The orange dot is the prediction provided by the LSTM RNN 1 model, based on the previous five InSAR PSI observation (i.e., yellow dots at times [-4, 0]). The grey dot is the prediction provided by LSTM RNN 2 model, based on the previous five InSAR PSI observation (i.e., yellow dots at times [-4, 0]), the past rain data (i.e., blue dots at time [-4, 0]) and the rain forecasting (i.e., blue dot at time 1).

5. CONCLUSIONS

This paper presented a comparison between two LSTM RNN. A traditional LSTM RNN 1 trained with InSAR data and able to predict future movements has been reproduced and used as term of comparison to evaluate the new approach proposed for LSTM RNN 2; the latter, was designed to combine InSAR data with meteorological data of the area of interest in the same period considered for the SAR acquisitions and the rain forecasting for the next acquisition. The paper applied this methodology to forecast the displacements of areas at risk of mudslides, which are triggered by large rainfalls. After training and testing of some parametric versions of the two proposed LSTM RNN, they were compared in terms of ability to predict future ground motion through MSE and MAE indices. The results show how the fusion of meteorological and InSAR PSI data improves the forecasting capacity of the measurement compared to the only usage of InSAR PSI data of LSTM RNN 1. Improvement is expressed in terms of reduction of both MSE and MAE indicators.

One of the possible future works is to understand the sensibility of the InSAR PSI predictions with respect to the variation of the forecasting rain data. Moreover, though the rain data are available daily, the InSAR PSI data remains with a quite small acquisition rate, new incoming big SAR constellations could be open to the possibility to improve the results of this work. Another future development consists in the adaptation of the proposed method to other applications than mudslides, by choosing the most suitable data to merge with the InSAR data to improve the predictions of InSAR PSI measurements. An example is the movements of road infrastructure, in which the InSAR PSI data can be used together

with traffic data for improving the ability of the LSTM RNN of making predictions.

REFERENCES

Agrawal, A., 2022: InSAR Data Analysis using Deep Neural Networks, IOP Conf. Ser.: Earth Environ. Sci. 1032 012025.

Blasco, J., Foumelis, M., 2018: SNAP as InSAR Processor For PSI Processing With StaMPS, Manual.

Chen, C., Song, S., Chen, J., Zheng, L., Kong, Y., 2017: An Approach to Predict Debris Flow Average Velocity, *Water* 9, no. 3: 205. doi.org/10.3390/w9030205.

Chen, Y., He, Y., Zhang, L., Chen, Y., Pu, H., Chen, B., Gao, L., 2021: Prediction of InSAR deformation time-series using a long short-term memory neural network, *International Journal of Remote Sensing*, 42:18, 6919-6942, doi.org/10.1080/01431161.2021.1947540.

Ferretti, A., Prati, C., Rocca, F., 2001: Permanent scatterers in SAR interferometry, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 8-20, Jan. 2001. doi.org/10.1109/36.898661.

Fiorentini, N., Maboudi, M., Leandri, P., Losa, M., Gerke, M., 2020: Surface Motion Prediction and Mapping for Road Infrastructures Management by PS-InSAR Measurements and Machine Learning Algorithms, *Remote Sens.* 12(23), 3976. doi.org/10.3390/rs12233976.

Hill, P., Biggs, J., Ponce-López, V., Bull, D., 2021: Time-Series Prediction Approaches to Forecasting Deformation in Sentinel-1 InSAR Data, *JGR Solid Earth*.

Hooper, A., Bekaert, D., Hussain, E., Spaans, K., 2018: StaMPS/MTI Manual Version 4.1b, University of Leeds.

Marinkovic, P., Ketelaar, G., van Leijen, F., Hanssen, R., 2007: InSAR quality control: Analysis of five years of corner reflector time series, *Workshop ESA SP Frascati, Italy* (pp. 26-30).

Naghbi, S., Khodaei, B., Hashemi, H., 2022: An integrated InSAR-machine learning approach for ground deformation rate modeling in arid areas, *Journal of Hydrology*, Volume 608, 2022, 127627, ISSN 0022-1694. doi.org/10.1016/j.jhydrol.2022.127627.

Zhang, Y., Hashemi, H., 2022: InSAR-Deep learning approach for simulation and prediction of land subsidence in arid regions, *EGU22*.