Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 2079–2090
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.7161

Seismic data analysis using feed forward BP neural network model for earthquake prediction

Oday Jalil Abdulelah, Sepanta Naimi*

Civil Engineering, Altinbaş University, Turkey

(Communicated by Javad Vahidi)

Abstract

Earthquakes are one of the most devastating and costly natural risks that a country faces, as they occur without notice and can result in major injuries or the loss of human lives as a result of damage to the destruction of a large number of houses, buildings, and other rigid structures. The point of this review is to assess the exhibition of Artificial Intelligence strategy in foreseeing the following event tremor utilizing the seismic wave signals. We present a three-layer feed forward BP neural organization model to find factors related to quake greatness M and seven other numerically determined boundaries. As info and target vectors, seismicity markers are utilized.

Keywords: Earthquakes, Artificial Neural Network (ANN), Seismic Forecasting
2020 MSC: 68T07, 68T09

1 Introduction

Since 1970 a generous progress in the applications of Intelligence Techniques like, linear programming, nonlinear programming and dynamic programming in natural hazards, geosciences and geotechnical engineering problems like rainfall analysis, ground water level changes, slope stability analysis, satellite image processing, petroleum exploration and estimation, Traffic Engineering, wind power forecasting, etc have been made. Recently evolutionary intelligent techniques like data mining, genetic algorithm, artificial neural networks and fuzzy logic gaining predictable result in various disciplines of sciences and engineering due to their effectiveness in tackling complex problems [1]. Notwithstanding, the use of these methods in geotechnical designing is new and restricted. Achievement of these techniques relies generally upon the picked hypothetical model for the framework to be dissected coordinating with the information yield information. In any case, there are circumstances where the arrangement is information driven, instead of demonstrating focused [13].

Machine learning and deep learning models are used in several domains and application. To address such issues, a new technique, Artificial Neural Network (ANN) based on Artificial Intelligence (AI), has recently been developed. Within a short amount of time, it was discovered to have broad applicability spanning multiple disciplines. As a result, there has been a rise in research effort in the art of applying such methods to real-world challenges, exposing the inherent potential and limitations of such systems. In general, the efficiency of all numerical approaches varies according to the problem [2]. As a result, none of these strategies can be used to solve all types of problems; there are various challenges that must be solved before alternative algorithms may be successfully applied.

*Corresponding author
Email addresses: 203724542@ogr.altinbas.edu.tr (Oday Jalil Abdulelah), sepanta.naimi@altinbas.edu.tr (Sepanta Naimi)

Received: July 2022 *Accepted:* October 2022

Many papers discussing the possible signals of different phenomena like Climate changes (Humidity, temperature, Rainfall), Ground water level changes, Electrical Current, Magnetic field, ULF-(Ultra Low Frequency) emissions, Animal behaviors, Frequently recorded seismic data base and GPS Method. Joseph LK (2000) evaluated a concise audit of conceivable seismic forerunners and presumed that slant, hygroreceptor (moistness), electric and attractive tangible frameworks in creature conduct, tactile physiology, and hereditary qualities could be connected into a seismic break social framework. He additionally referenced seismic quiet before enormous consequential convulsions, radon lessening, and change in ground water level [3].

1.1 Seismology

Earthquakes (EQ) are quite possibly the most ruinous and exorbitant regular hazard that a nation faces, as they happen abruptly and can bring about genuine wounds or the deficiency of living souls because of harm to and annihilation of an enormous number of properties, structures, and other unbending constructions [12].

A quake is characterized as a speedy shift of the world's surface brought about by the arrival of energy from the world's hull. The tremor expectation is to distinguish a positive date, area and extent for a quake inside expressed vulnerability impediments. The motivation behind seismic tremor forecast is to give early notice of possibly harming quakes, permitting individuals to react suitably to the debacle and limit death toll and property. Many examinations have been led to explore the potential indications of different peculiarities like seismicity, power, and radiance that either go with or are trailed by tremors [8]. The idea that there should be exactly discernible archetypes to seismic tremors is naturally enticing, yet examinations over the past 120 years have neglected to affirm it, hence individual quakes are eccentric. The overall agreement on the subject of seismic tremor expectation is that the earth is a complex nonlinear framework, as shown in Figure 1.

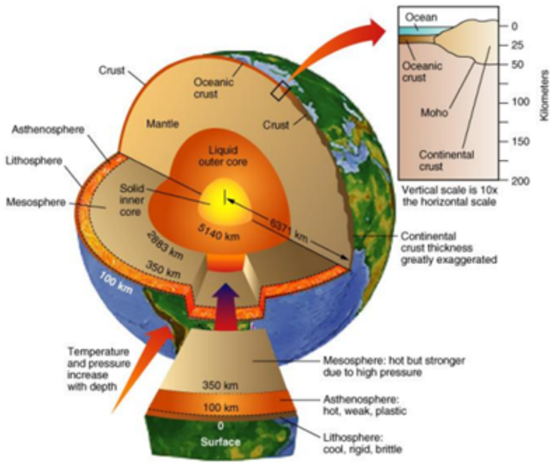

Figure 1: Earth's internal structures

The covering can be seismically actuated with a generally little irritation of the general driving conditions, and convoluted non-direct seismology does not have an all around concurred actual model; the arbitrariness in the inception of an enormous crack thwarts the utilization of short expectation windows; quake event is viably stochastic, and endeavors to accomplish deterministic forecast seem unjustifiable; blames and issue frameworks are inhomogeneous; seismicity at a shortcoming framework is inhomogeneous; Earthquakes positively react to pressure changes from past tremors, however the reaction is perplexing [9]. Exceptionally huge tremors happen too rarely to test theories on the best way to foresee them with the factual meticulousness one might want; there is still no acceptable hypothesis about the nucleation of quakes; there is just a simple comprehension of the physical science of seismic tremor bursts; anticipating tremors requires a comprehension of the fundamental physical science; dependable forerunners are hard to decide, yet they are probably not going to exist on exact actual grounds [4].

The presentation of broadband seismometers during the last decade has given great and definite accounts of ground movement during a few huge quakes and little vibrations too.

1.2 Problem Description

Among all natural disasters, earthquakes pose the most serious threat to humanity. They occur frequently and without warning. Although earthquakes are unavoidable, a complete assessment of seismic hazard and risk reduction can help to reduce social and economic costs after an earthquake. During the twentieth century, big earthquakes killed approximately 17,000 people each year on average. The majority of these earthquakes occurred along tectonic belts of plate borders all around the planet, including the circum-Pacific ring of fire. Because of the severity of severe earthquakes and the immense damage they cause, there is always an urgent need for earthquake predicting studies, construction rules, also, safe constructions, especially in emerging nations like India. The progressions in quake hazard evaluation and danger the board that follow will bring about gigantic reserve funds in human existence and property.

Real-time seismology, which deals with the difficulty of estimating the magnitude and position of an earthquake from the beginning of the rupture process, provides the scientific foundation for the notion of early-warning relays. As a result, earthquake prediction is usually regarded as one of the most difficult scientific topics, owing to both its societal importance and the inherent complexity of the subject. Earthquake prediction, by providing basic information about expected earthquake times and locations, might aid in reorienting present strategies toward enhanced earthquake preparedness.

1.3 Motivation

Earthquakes are one of the most expensive natural risks that the country faces since they strike without notice and can result in significant injuries or the loss of human life as a result of damage to buildings or other inflexible structures. The weakness of mankind to catastrophes brought about by the development and centralization of populaces, economies, radioactive, harmful, and other hazardous materials and ventures has expanded significantly over the most recent couple of many years and keeps on rising; many serious quakes have brought about billion-dollar property misfortunes. While seismic tremors are by all account not the only wellspring of risk, six of them are positioned as among the world's main 20 catastrophic events somewhat recently of the 20th century. These quakes were a long way from the best, however they represented 35% of absolute monetary harms from cataclysmic events, in front of floods (30%), windstorms (28%), and others (7%), as indicated by CAT-I administration information (2000). Seismic tremors are additionally the most deadly of every normal catastrophe. To effectively alleviate tremor harm, endeavors should move away from exceptionally costly post-catastrophe salvage activities and toward information based danger strong public resources. There are around 80,000 human deaths every year, with 220 occurring per day. The table below displays the average number of earthquakes registered per day and year.

2 Proposed System

ANNs are grouped into three types based on their architecture: Back Propagation Neural Networks (BPNN), category learning (unsupervised) networks (self organizing map), and probabilistic neural networks. BPNNs are more suited for prediction problems, whereas categorical learning ANNs are commonly employed for classification problems. Back propagation employs gradient descent rules, category learning employs Kohonen learning laws, and Probabilistic Neural Network (PNN) employs both Kohonen and probabilistic learning laws. Another type of BPNN is the recurrent neural network, and probabilistic neural networks include the Generalized Regression Neural Network (GRNN) and the Radial Basis Neural Network (RBNN).

One of the most significant and hardest aspects in the model-building process is determining optimal network architecture (geometry). Nodes in one layer are only connected to nodes in the following layer in feed forward BPNN. In recurrent networks, however, nodes in one layer can be linked to nodes in the next layer, the previous layer, or the same layer. BPNN is the most commonly used architecture in geotechnical engineering; other networks, such as recurrent networks, are utilized for modeling stress-strain characteristics. For pile capacity, use GRNN. RBNN for site characterization and probabilistic neural network for liquefaction analysis (Shahin et al. 2004) have used unsupervised learning network (self-organizing map) for input data clustering while applying BPNN for prediction of settlement of shallow foundation.

In the present work, strong motion instrument data from LANL earthquake dataset, are examined, in all, 462 records analyses with magnitude varying from 3.0 to 7.0 and epicentral distance in the range from 5 km to 370 km are now available. These have been analyzed to understand the characteristics of earthquake ground motion. The characteristics studied are the peak parameters, occurrence time, regions, latitude, longitude, duration of the strong shaking and ductility reduction factor. Results obtained are compared with those from newly recorded year seismic data for the Himalayan regions based on earthquakes elsewhere and with currently adopted design practices. Other than earthquake prediction helps the people life from the followings.

1. Nuclear power plant
2. Chemical & Fertilizers making industries

3. Mining 4. Thermoelectric power plant
5. Railways
6. Large building collusions
7. Oil, Petroleum and natural gas eruptions
8. Electric Grid transmission and distributions
9. Tsunami
10. Wind power

2.1 The Basic Principles and Methods of BP Neural Network

A neural organization is a major organization set up by interfacing an enormous number of essential neurons to shape a complicated organization, with an equal appropriated data handling design of the compound by a non-straight capacity to surmised the planning between the information and result. It utilizes simple neurons, memory, and relationship to manage a wide scope of questionable non-straight information, deciding, computing, and dissecting it utilizing a versatile example acknowledgment approach. As portrayed in Figure 1, the neuron is the essential taking care of unit of neural associations, with n inputs; every data is taken care of by a fitting weight w related with the neuron, and its criticism yield relationship can be depicted as follows:

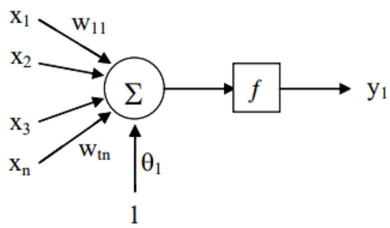

Figure 2: Basic neuron structures

$$I_i = \sum_{j=1}^{n} w_{\mathrm{ij}} x_j - \theta_i \tag{2.1}$$

$$y_i = f\left(I_i\right) \tag{2.2}$$

where $x_j \quad (j = 1, 2, 3......., n)$ is the quantity of neurons in the information signal, I is the nerve cell edge, w_{ij} is the association loads from neuron j to neuron I, and n is the quantity of neurons in the information signal. The neuron yield is signified by yi, while the result change work is meant by $f()$. The information (pointers) contribution to the information layer of every hub to move from the info signal communicated to every hub in the secret layer is the BP calculation.

After the role of transfer function [Sigmoid Function: $f(x) = 1/(1 + e - x)$] communicated the result sign of each secret hub to the result hub, at long last the result signal arrived at the result layer a result. In case the result layer can't get the ideal result signal, then, at that point, the blunder of the result signal from the result layer through secret layer to include layer back-spread to change the neural association loads and the edge worth of the neuron move work thus that the deviation between the model result and the ideal result is situated inside the arrangements.

2.2 BPNN Network Description

Artificial neural networks are fully used in the forms of feed forward supervised neural networks with error back propagation learning method (BP). This networks are called back propagation networks (BP ANNs). The learning method is the gradient descent method. BPANNs have two major advantages.

1. The learning system can be accelerate with the assistance of techniques for numerical programming;

- Conjugate gradient methods (CG)
- Quasi network methods (QN)

When compared to QN, CG methods are faster and need less memory, but they are less trustworthy. The QN approaches are the fastest, but they include matrices, which requires a lot of memory, and they have nothing in common with the learning process in biological networks. Experiments reveal that CG and, in particular, QN approaches are at least 5 times faster and more reliable than the best BP methods.

2. Finding an optimal configuration is most difficult one. The following parameters of the ANN should be set:

- The number of hidden layers,
- The number of neurons in the each layers
- The sort of the actuation work which can fluctuate in a similar layer.

These boundaries should be painstakingly picked to keep away from overfitting or underfitting and to make the learning stage ideally focalized. According to the Kolmogorov theorem, the majority of BPANNs have two hidden layers. More hidden layers rarely increase performance. Some basic rules exist to define the number of neurons in hidden layers. The previously covered up layer ought to have twice however many neurons as there are input neurons. The quantity of neurons in the following secret layer ought to be somewhere between the quantity of neurons in the result layer and the quantity of neurons in the principal stowed away layer.

2.2.1 The Multi-layer BP network design issues to be considered

Making the BP network setup ought to think about the amount of layers, the amount of neurons in each layer of the association, the fundamental worth, and the learning rate.

1. The amount of association layers displayed by a Multilayer BP association. The greatest number of layers can additionally diminish blunders by expanding preparing time and diminishing mistake. Preparing is simpler with a bigger number of neurons in the covered layer.
2. To improve accuracy, the number of hidden layer neurons must be maximized, as well as the number of output layers with linear activation functions.
3. The starting worth of the right worth is picked. The Back spread organization is a very successful device for producing a non-direct exchange work between various nonstop esteemed sources of info and at least one persistent esteemed results.
4. Learning rate: There is an appropriate learning rate for every novel organization, however for more complicated organizations, the various parts of the mistake complex organizations are available. To decrease preparing times, the learning rate and preparing time still up in the air. The expected mistake is: By looking at the minima, the secret layer hubs, the anticipated blunder not set in stone. Because of an examination of two distinctive anticipated mistakes of the organization, a secret layer of the neural organization is prepared to accomplish any persistent capacity estimate.

2.3 The BPNN Learning Algorithm

A multi-facet perceptron is a feed-forward neural organization (ANN) that looks to build an ideal connection in an information/yield set of learning designs. An information vector layer, various secret layers, and result vector layers make up a neural organization. Each layer contains its own arrangement of neurons, each with its own arrangement of weight associations. A solitary preparing design is an Equation 2.3, 2.4, 2.5 I/O vector of sets of information yield esteems in the whole lattice of I/O preparing set. The neural organization model is displayed in Figure 7);

The information $x_i, i = 1, 2, ...n$ that the info layer gets is comparable to the electrical sign got by neurons in the human cerebrum. These info signals are duplicated by association loads $w_{p,ij}$ in the least difficult model, and the successful information $\mathrm{net}_{p,j}$ to neurons rises to the weighted amount of the data sources.

$$\mathrm{net}_{p,j} = \sum_{i=1}^{n} W_{p,ij}\mathrm{net}_{q,i} \tag{2.3}$$

$W_{p,ij}$is the interface weight of the layer p from the i neuron in the q (source) layer to the j neuron in the p (target) layer, $\mathrm{net}_{q,i}$ is the result provided at the i neuron in the layer q, and $\mathrm{net}_{p,j}$ is the result created at the j neuron in the layer p. For the information layer, inputs x_i are compared to $\mathrm{net}_{p,j}$.

The processed output(s), otherwise called the noticed output(s), are deducted from the ideal or target output(s) at the result layer to create the blunder signal.

$$\varepsilon(W) = \frac{1}{2m}\|E(w)\|^2 \tag{2.4}$$

$$E_i(W) = \sum_{j=1}^{t}[\mathrm{out}_{k,j} - \mathrm{tar}_{k,j}] \tag{2.5}$$

where m denotes the number of prepared sets, tark ,i, and out k,i denote the desired and observed output(s) for hub i in yield layer k, respectively. This method of preparing for ANN is known as controlled learning..

A learning calculation endeavors to assess the loads to get the most ideal reaction for each information vector took care of into the organization. Iteratively, the mathematical minimization calculations utilized for preparing create a grouping of weight networks. To utilize an algorithmic activity A., a beginning worth of the weight grid w(0) Equation 2.6 is required, while the emphasis recipe is as per the following:

$$W^{(t+1)} = A\left(W^{(t)}\right) = W^{(t)} + \Delta W^{(t)} \tag{2.6}$$

The aforementioned formula serves as the foundation for all numerical methods used in ANNs. The algorithm's altering component is further subdivided into two pieces as Equation 2.6

Training methods are classified into two types. Worldwide methodologies are calculations that influence worldwide information on the condition of the whole organization, for example, the course of the general weight update vector. Nearby transformation methods, then again, depend entirely on weight-explicit data, like the worldly conduct of the incomplete subordinate of the weight. The neighborhood approach is all the more straightforwardly attached to the ANN thought of disseminated handling, which permits calculations to be made free of each other. Besides, apparently, despite the fact that utilizing less data, nearby strategies give quicker and more dependable expectation than worldwide techniques by and large. To get the important association loads from the arrangement of P input-yield designs in the preparation set, the learning technique goes as follows.

Step 1: Initialize the connection weights w_{ij}

Step 2: Initialize the iteration count I

Step 3: Read in an input pattern p

Step 4: Starting from the nodes in the first hidden layer, compute the Neuron output using Equation 2.6

Step 6: Repeat steps 3 to 5 until all patterns in the training set have been considered (p = P)

Step 7: Compute the error function E(I) for the present iteration (iteration I) using the following equation

Back propagation is a method that searches for a solution in continuous space. As a result, when used to a binary-to-binary mapping problem, the Back propagation approach may result in a lengthy training period and inefficient performance. To obtain even a simple binary-to-binary mapping, the BPA typically requires an incredibly large number of iterations.

2.4 Seismic Forecasting Model Based on BP Neural Network

The main advance in planning a Backpropagation counterfeit neural organization model is to separate pertinent component vectors that are a sensible impression of the model's feedback and result rationale of correspondence between. The info vector and result vector are both remembered for the element vector. The decision of info vector expansion in the neural organization model of seismic estimating considers not just the seismic effects of encompassing effect, but also the ease of energy, latitude, longitude, time, magnitude, regions under different date.

According to seismic wave and energy release from wave calculate damage factors [14, 5]; use the following 9 earthquake computational parameters: epicentral latitude layers x_1, epicentral longitude x_2, Magnitude mean x_3, b values x_4, energy E x_5, energy J x_6, coefficient of variation x_7, Mean square deviation x_8, and class x_9, as the input vector component, the input vector $X = (x_1, x_2, \ldots, x_9)$.

The x boundaries, as information an element vector part of the administrative qualities, contingent upon the conditions and embracing various strategies, explicit boundary esteems are as per the following:

1. *b* - worth: The b-value is a measure of the frequency of mild to big earthquakes in a certain area over a specific time period. Numerous studies calculated the b-value in intraplate zones, as well as locally, regionally, and globally. With the exception of Asia, where the b-value spans from 0.90 to 2.1 at a corner magnitude Mc of 6.0 to 7.0, the low b-value is a feature of intraplate regions.
2. Mmean: Average value of last two earthquake magnitude
3. Energy E: Most calculations of the magnitude-energy relation depend directly or indirectly on wave group from a point source, (E in ergs)
4. Energy J: Seismic wave energy on surface
5. Latitude: Epicentral latitude 6. Longitude: Epicentral longitude
6. Class: characterization of seismic waves as per size.

Table 1: Architecture and training parameters of the proposed BPANN

Architecture	
Number of Layers	3
Number of neurons on the laters	input 9, Hidden 12, Output 1
Activative Function	Tan-Sigmoid, Pureline
Training Parameters	
Larninig rule	Levenberg-Marquardt
Adaptive Learning rate	Initial:0.01, increase: 1.04, decrease:0.4
Momentum constant	0.94
Sum-squared error	0.004
Epochs	10000

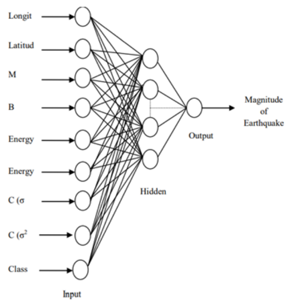

Figure 3: Neural network architecture that was used in modeling magnitude of earthquake

This study demonstrates neural network algorithms to predict the occurrence of next earthquake using the seismic wave data provided.

3 Results and Discussion

In this section we present the implementation of the proposed Earthquake Prediction using the proposed BPANN. The goal of this project is to predict upcoming earthquakes based on a stream of seismic activity. The training data consists of a ~10GB large csv file which contains a seismic_activity and a time_to_failure column, which represents how many seconds are left before the next earthquake:

3.1 DataSet

A large focus of Earth science research belongs to forecasting the timing and severity of earthquakes. Earthquakes are notoriously difficult to predict and can have devastating impact on those unfortunate enough to experience one. One of the few predictors of earthquakes known to scientists is overall seismic activity measured from vibrations caused by movement in tectonic plates in the earth's crust. The dataset used in this research is LANL Earthquake-Prediction dataset. The data comes from a well-known experimental setup used to study earthquake physics. The acoustic_data input signal is used to predict the time remaining before the next laboratory earthquake (time_to_failure). The training data consists of $a \sim 10GB$ large csv file which contains a seismic_activity and a time_to_failure column, which represents how many seconds are left before the next earthquake. Our goal is to forecast when the next earthquake will take place based on seismic data using time series analysis and machine learning techniques.

Table 2: The training data

Sesmic activity (v)	**Time to failure (s)**
12	1.4690999832
6	1.4690999821
8	1.469099981
5	1.4690999799
8	1.4690999788
8	1.4690999777
9	1.4690999766
7	1.4690999755
-5	1.4690999744
...	...

3.2 Feature selection

We ended up with over 2,000 features in our feature store. Some of these turned out to be identical, almost perfectly correlated, or constant, and so were removed. The hard part was figuring out what to do with many basic features (such as the mean) whose distribution in the test set was very different than on the train set:

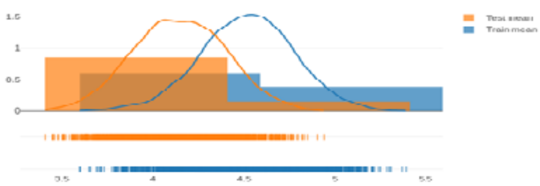

Figure 4: Data Distribution

3.3 Data Exploration and Exploratory Visualization

The dataset is generated from the model laboratory earthquake setup depicted in Figure 5 [3]. In reality, earthquakes occur when two tectonic plates move against each other. The rough interfaces get stuck against each other and start storing energy like a spring. Eventually the friction between the plates becomes unsustainable and the stored energy is released, causing vibrations to spread through the plates. The y set-up models this behavior by forcing a plate to move past two adjacent frictional interfaces. Similar to real earthquakes, the plate will periodically get stuck and then release energy, resulting in a 'labquake'. This can be detected by a sensor that tracks the frictional force in the contact interface. A separate piezoelectric sensor is attached to one of the plates and tracks the vibrations at a sample rate of 4MHz.

The used dataset provided on the Kaggle website [3] consists of a 9+Gb csv file with two fields: 'acoustic_data' and 'time_to_failure'. The data represents approximately 158s of a continuous lab earthquake experiment. Every 1000th row of the dataset is sampled and plotted in Figure 6: Training dataset sampled at every 1000th row in order to visualize the entire <<dataset. The key global statistics of the data are subsequently presented in Table 3. Per the table, there are a total of 629.1 million rows in the dataset.

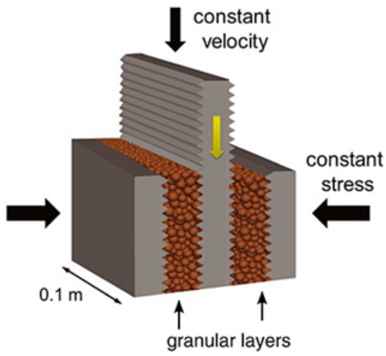

Figure 5: earthquake test set-up [3]

Table 3: Global statistics of dataset

	acoustic - data	**time-to-failure**
count	6.291455e+08	6.291455e+08
mean	4.519468e+00	5.678292e+00
min	-5.515000e+03	9.550396e-01
max	5.444000e+03	1.610740e+01
std	1.073571e+01	3.672697e+00

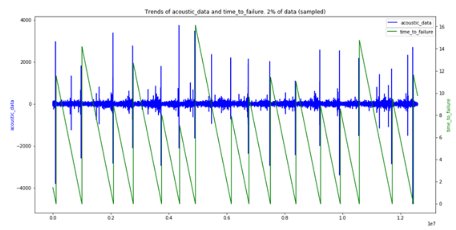

Figure 6: Training dataset sampled at every 1000th row

The green plot in Figure 6: Training dataset sampled at every 1000th row, representing the 'time_to_failure' values, can be seen to hit 0 before spiking multiple times. These represent 16 separate labquakes that occur within the experiment data. The 'time_to_failure' values are the regression targets for the model. The blue plot represents the 'acoustic_data'. Per the work details, each input 'datapoint' for the solution BP ANN model is a 0.0375s segment of the 'acoustic_data' (150,000 rows). An example of a datapoint is presented in Figure 7.

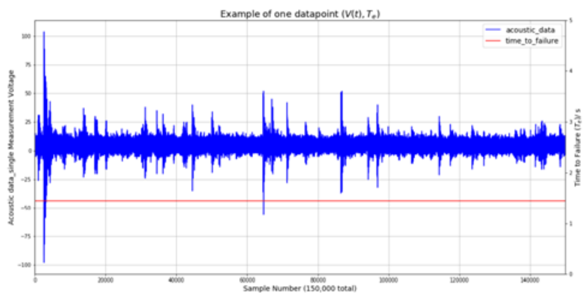

Figure 7: Example of 1 'datapoint'

Exploring the dataset, the following preprocessing observations are made:

1. The input dataset is a stream of continuous raw data. It will have to be converted into a set of input/output datapoints such as the example in Figure 7. Each input is a 0.0375s (150,000 rows) segment of the 'acoustic_data'. The target is a single 'time_to_failure' value from the last row in the 0.0375s segment.

2. The raw data contains 16 labquake spikes. These regions must be excluded when preprocessing the input/output datapoints
3. The input datapoint consists of 150,000 'acoustic_data' values. This is obviously an extremely high number of dimensions. Training a model with the raw values would be computationally expensive, and probably contains excessive amounts of noise. Feature engineering will therefore be a critical component of the path to the optimal solution. The input will have to be processed into a set of features that capture the critical aspects of the 0.0375s time series.

3.4 Benchmark

A benchmark model will be made by implementing the machine learning algorithm suggested in previous section. The prior research describes a BP ANN model trained on simplistic statistic features such as the mean and standard deviation of the input time-series. The output of the model is the estimated time_to_failure. This model represents the current state of the art. It can be recreated for the current dataset, and used as a benchmark with which to compare future models. The comparison will be quantified with the mean absolute error calculated on cross-validation or test sets.

Using the most basic set of features, an optimized BP ANN model benchmark results in a mean absolute error (MAE) of 2.28 on a test set made from 35% of the input data. The BP ANN model was optimized with the GridSearchCV class in sklearn [4]. The final hyperparameters used for the benchmark were: max_depth = 5; min_samples_leaf = 0.05; n_estimators = 30. The error criterion is equal to the MAE. The benchmark prediction can be visualized against the full set of available data in Figure 8. The implementation is detailed in the 'Benchmark Model_fset1.ipynb' notebook. The comparison shows that the benchmark is able to roughly predict the trend of the time_to_failure values, but fails at predicting the

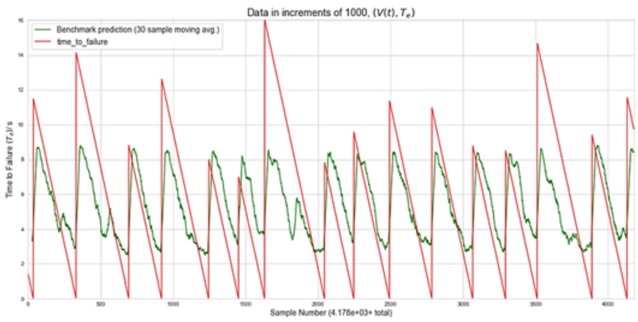

Figure 8: Actual time_to_failure vs Benchmark prediction

3.5 Model Evaluation and Validation

The final model is a BP ANN trained on all of the developed features. This model lead to the highest MAE score in the test set (2.029). The robustness of the model is proven by a 5-fold cross-validation analysis. The final mean MAE score was 2.040, with a standard deviation of 0.045. The fact that the standard deviation was under 3% of the mean, shows that despite perturbations in the input data, the chosen model produced similar results.

The 'time_to_failure' prediction results from the final model (blue), the benchmark (green), and the actual values (red) are plotted and compared in Figure 9. The benchmark was earlier stated to capture the general trend of the 'time_to_failure' values, but not able to capture the extremes. The plots in Figure 9 show that the final model is consistently between the green benchmark and red target values for 'time_to_failure'. The final model is therefore definitely better at predicting the time until the next labquake than the benchmark. The predictions capture the general trend, but are also closer to the extremes. Nevertheless, the final solution is still not able to capture most of the extremes from the 'time_to_failure' values. For example, the blue lines never go below 1.5s in the 'time_to_failure' plots. The achieved MAE score on the unknown earthquake data is 1.630, which improves on the 1.734 achieved by the benchmark. For comparison, the current top score on the leaderboard is 1.282. Overall, the solution is not considered significant enough to have 'solved' the problem. Further feature exploration and model tuning will be required to improve model performance.

Figure 9 presents a comparison of the benchmark, final model, and actual data values for the time remaining until the next labquake in the provided data. As reported in the Results subsection, the final solution performs better than the benchmark, but is still unable to adequately capture the labquake phenomena. A further aim of the project was to identify the key feature importance's that could be used to predict earthquakes.

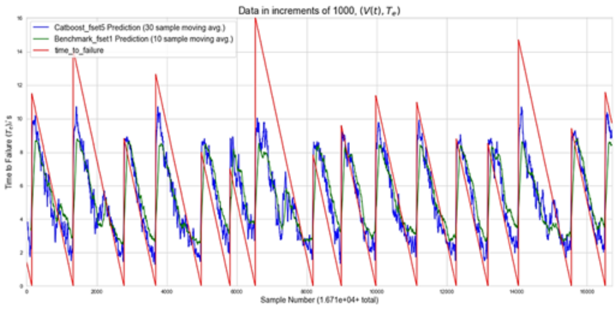

Figure 9: Comparison of Final Model, Benchmark, and actual time_to_failure values

For feature engineering, additional methods to explore include wavelet decomposition, and possibly some form of principal component analysis. Both of these methods have been successfully used on time-series before. In addition, the final model in the project implementation used all the explored features. However, to make the solution more efficient, features with low importance could be trimmed. This would reduce the 'noise' from unimportant features affecting the model and causing overfitting. For example, the sklearn feature engineering package could be used.

3.6 Computational and Statistical Analysis

The computed values of BP ANN trained and predicted values in our dataset were compared to the original seismometer recorded values. Table 4 displays the Error % of successful and unsuccessful prediction ratios, as well as a comparison chart.

Table 4: Comparison between computed values of BPANN predicted values and originally recorded values

Input class	**Output Magnitude Ranges**	**BP ANN Predicted Values**	**Originallay Recorder Values**	**Success ranting In %**
Class 1	< 3.0	2	3	66.66 %
Class 2	3.0 – 3.4	4	3	133 %
Class 3	3.4 – 3.8	3.2	3.5	91.42 %
Class 4	3.8 – 4.2	3	4	75.00 %
Class 4	3.8 – 4.2	5	4	125 %
Class 5	4.2 – 4.6	4	4.5	88.88 %
Class 6	4.6 – 5.0	3.8	4.7	80.85 %
Class 7	5.0 – 5.4	5	5.5	90.90 %
Class 8	5.4 – 5.8	4.7	5.9	79.66 %
Class 9	5.8 – 6.2	0	0	0 %
Class 10	6.2 – 6.6	0	6.35	0 %
Class 11	6.6 – 7.0	0	0	0 %
Class 12	7.0 – 7.4	0	7.25	0 %
Class 13	7.4 – 7.8	0	0	0 %
Class 14	7.8 – 8.2	0	0	0 %
Class 15	8.2 – 8.6	0	0	0 %

4 Conclusion

The application of soft computing tools like Artificial neural network, Association rules, evolutionary algorithms, genetic algorithm and fuzzy logic in different branches of science and engineering discipline is phenomenal. However, the applications of the above techniques in geotechnical engineering are very limited. In this study artificial neural network model based on Backpropagation neural network developed and utilized for the seismic data analysis

References

[1] A.S.N. Alarifi, N.S.N. Alarifi and S. Al-Humidan, *Earthquakes magnitude predication using artificial neural network in northern red sea area*, J. King Saud Univ. Sci. **24** (2011), no. 4, 301–313.

[2] H. Adeli *Neural networks in civil engineering: 1989-2000*, Comput. Aid. Civil Infrast. Engin. **16** (2000), no. 2, 126–142.

[3] H. Adeli and S.L. Hung, *Machine Learning: Neural Networks, Genetic Algoritms and Fuzzy Systems*, Wiley, New York, 1995.

[4] A. Paul, A.R. Prasath and R. Singh , *Slip heterogeneities evaluated for earthquakes* $M > 4.0$ *using waveform modeling in the garhwal region of central seismic gap in Northwest Himalaya, India*, Himalayan Geo. **36** (2015), no. 2, 153–160.

[5] W. Alfaggi and S. Naimi, *An optimal cost estimation practices of fuzzy AHP for building construction projects in Libya*, Civil Engin. J. **8** (2022), no. 6, 1194–1204.

[6] G. Asencio-Cortés, F. Martínez-Álvarez, A.Troncoso and A. Morales-Esteban, *Medium-large earthquake magnitude prediction in tokyo with artificial neural networks*, Neural Comput. Appl. **15** (2015), 2121–2127.

[7] D. Yang and K. Yang, *Multi-step prediction of strong earthquake ground motions and seismic responses of SDOF systems based on EMD-ELM method*, Soil Dyn. Earthquake Engin. **85** (2016), 117–129.

[8] E. Florido, F. Martínez-Álvarez, A. Morales-Esteban, J. Reyes and J.L. Aznarte-Mellado, *Detecting precursory patterns to enhance earthquake prediction in Chile*, Comput. Geosci. **76** (2015), 112–120.

[9] H. Bhaumik, S. Bhattacharyya, M. Das Nath and S. Chakrabory , *Hybrid soft computing approaches to content based video retrieval: A brief review*, Appl. Soft Comput. **46** (2016), 1008–1029.

[10] J. Geng, Y. Bock, D. Melgar, B.W. Crowell and J.S. Haase, *A new seismogeodetic approach applied to GPS and accelerometer observations of the 2012 brawley seismic swarm: Implications for earthquake early earning*, Geochem. Geophys. Geosyst. **14** (2013), 2124–2142.

[11] M.K. Kachakhidze, N.K. Kachakhidze and T.D. Kaladze, *A model of the generation of electromagnetic emissions detected prior to earthquakes*, Phys. Chem. Earth **85-86** (2015), 78–81.

[12] C. Marco-Detchart, J. Cerron, L. De Miguel, C. Lopez-Molina, H. Bustince and M. Galar, *A framework for radial data comparison and its application to fingerprint analysis*, Appl. Soft Comput. **46** (2016), 246–259.

[13] S. Narayanakumar and K. Raja, *A BP artificial neural network model for earthquake magnitude prediction in himalayas, india*, Circ. Syst. **7** (2016), 3456–3468.

[14] S. Moazam, B. Boroomand, S. Naimi and M. Celikag, *Wave propagation in unbounded domains under a dirac delta function with FPM*, Math. Prob. Engin. **2014** (2014).

[15] U. Raghavendra, U.R. Acharya, H. Fujita, A. Gudigar, J.H. Tan and S. Chokkadi, *Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images*, Appl. Soft Comput. **46** (2016), 151–161.