Deliverable 4.3

Machine learning approaches for earthquake detection

Organisation(s)	University of Oxford, Equinor, NORSAR
Author(s)	Tom Kettlety Tesfahiwet Abraha J Michael Kendall
Reviewer	Anne-Kari Furre Nadege Langet
Type of deliverable	Report
Dissemination level	Open
WP	4
Issue date	20 February 2024
Document version	1

Keywords:

Seismology, earthquake detection, machine learning, deep learning, phase picking.

Summary:

This report is a review of machine learning-based methods for earthquake detection. It first introduces the basic concepts of traditional earthquake detection, and commonly used approaches. It then goes onto describe the concept of artificial intelligence, machine learning, and deep learning. This is followed with a summary of how these concepts are employed to detect earthquakes in seismic data, with a synthesis of the most widely used machine learning detection algorithms. The ways in which machine learning methods could be applied to the activities in the SHARP project are given in the conclusions, along with the future directions of the field.

SHARP storage project Deliverable 4.3: Machine learning approaches for microseismic detection

February 2024

Contents

Executive summary

This report provides a review of the application of artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques to earthquake detection. We start by reviewing the conventional methods for earthquake detection. We then review the fundamental concepts behind artificial intelligence, machine learning, and deep learning. We then describe how these new techniques are being employed in earthquake detection, and explain how the current deep learning architectures and models are being applied. We conclude with a discussion on how ML and DL methods can be effectively applied to geological CO_2 storage monitoring.

The report will present an overview of the existing landscape but also summarise the current challenges and future directions of machine learning in its application to earthquake detection. By addressing the opportunities and limitations, this report aims to contribute to a deeper understanding of this oft-invoked field, and how these methods could improve earthquake detection and potentially offer innovative solutions in CO_2 storage applications.

AI, and more specifically ML, have a wide variety of applications, and offer the ability to automate algorithms and computational processes that can otherwise be time consuming and labor-intensive. Earthquake detection is a clear example of one such process. Accurate and reliable routines for detecting earthquake arrivals from continuous waveform data are essential for all seismological analyses. Thus, finding methods to improve the detection capability for events is key to improving our understanding of the subsurface that stems from the analysis of earthquakes.

Many ML models have been developed for earthquake detection and location. Each are trained on large datasets from specific regions, and thus have ingrained biases for the features of the underlying training data. Generally speaking, ML-based detection capability for a specific purpose can be improved by tuning the larger models using transfer learning, where the model is additionally trained on data from that specific network (with its own geometry, noise profile, event types, etc).

The performance of models can vary significantly. This is often considered to be due to features (network geometry, sensor type, station site conditions, event magnitudes, event distances, faulting style, etc) of the underlying data used to train the model. One clear drawback in the use of ML models, particularly for DL architectures, is their opacity with respect to feature extraction – it is often impossible to determine exactly what an ML model is identifying in the data. However, there are architectures which are less opaque, and can enable something approaching an understanding of the underlying features the ML algorithm is “seeing” in waveform data.

Many activities within the SHARP project could be enhanced through ML techniques. The earthquake catalogue data could naturally be enhanced through better detection methods, finding smaller and a more diverse range of events. ML filtering methods of the waveform data could be applied to improve signal quality determination and quality control methods for processing in workflows such as focal mechanism inversion or stress drop measurement. Whilst this report focuses on earthquake detection, other ML-based techniques are discussed. For example, image recognition methods could be employed to detect site amplification effects, which are key to accurately assessing ground motions and seismic hazard.

1 Introduction

1.1 Earthquake detection

Earthquake detection is a critical aspect of seismology and geophysics [Yoon et al., 2015]. It aims at identifying the occurrence of seismic events that occur beneath the Earth's surface. It involves the use of various ground motions sensors, such as seismometers, accelerometers, geophones, or fibre optic distributed acoustic sensing (DAS) cables to record vibrations caused by seismic sources. These recorded data can then be analysed to determine the location, depth, magnitude, and source properties of earthquakes. The process of earthquake detection typically begins with the collection of continuous seismic waveform data – showing the ground motion through time in one to three dimensions – from a network of seismic stations, which are then subjected to various signal processing and analysis techniques.

Accurate and reliable earthquake detection can be broken down into manual or automated methods. Manual picking involves experts reviewing seismograms to identify seismic events, while automatic detection uses various numerical or statistical techniques to identify signals of earthquake arrivals in continuous waveform data. A combination of the two are generally employed to ensure a robust earthquake monitoring system, where automatic pick times are reviewed by seismologists to improve their accuracy. Obtaining accurate estimates for the initiation time of detected seismic phases is a crucial first step in any seismological analysis [Leonard, 2000].

Manual picking of seismic events is an important and intricate process in seismology. Generally, experience is needed to identify the seismic phases correctly and to pick onset times accurately and consistently [Küperkoch et al., 2012]. Before the proliferation of computational methods, detailed studies of seismicity were built upon the foundation of manual picking. However, this is a time-consuming task [Baer and Kradolfer, 1987] and as the demand for real-time early warning systems and the analysis of larger data sets became more common, there has become an increasing reliance on automatic picking methods [Álvarez et al., 2013].

Numerous triggering algorithms can be employed to detect the onset of seismic arrivals in continuous data. These algorithms are generally based on the statistical characteristics of the signal in the time or frequency domain [Allen, 1978, 1982, Murdock and Hutt, 1983].

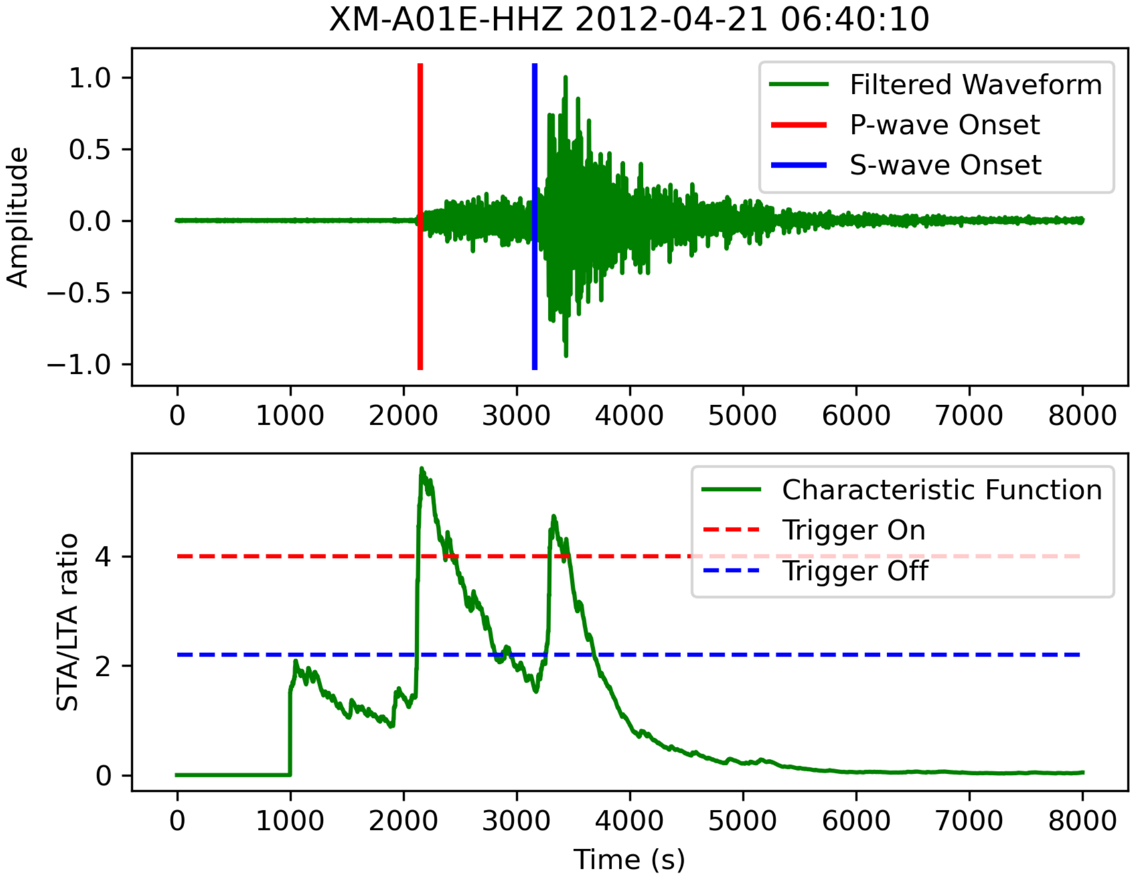

Figure 1: An example of short-term average (STA) and long-term average (LTA) with trigger on and off. The seismic waveform is recorded by station A01E near Aluto volcano in the Main Ethiopian Rift. The seismic event occurred on April 21, 2012, at 06:40:10 UTC. This waveform data is band-pass filtered between 2 Hz and 15 Hz. The upper plot shows the filtered waveform, highlighting the identified P-wave (red) and S-wave (blue) onsets. The lower subplot depicts the characteristic function (STA/LTA ratio), calculated using an STA window of 1.5 seconds and an LTA window of 10 seconds. The red dashed line in the bottom subplot represents the trigger-on at a threshold of 4.0, and the blue dashed line represents the trigger-off at a threshold at 2.2. These thresholds signify when the short-term variations significantly exceed the long-term trends.

The simplest of these rely on an amplitude threshold trigger, whereby a detection is made when some quantity based on the amplitude of the seismic signal surpasses a set threshold. A widely adopted algorithm in weak-motion seismology is the "short-time-average long-time-average trigger" (STA/LTA) [Trnkoczy, 2009]. This algorithm continually computes the absolute amplitude averages in short and long time windows, and triggers when their ratio surpasses a predefined threshold [Yoon et al., 2015]. This is depicted in Figure 1. Many trigger algorithms exist, including Joswig [1990], Joswig and Schulte-Theis [1993], and Joswig [1995]. In addition to the STA/LTA, there are other earthquake detection methods that are frequently used like template matching, auto-correlation, coalescence and migration.

Template matching (TM) involves comparing incoming seismic data to predefined templates of known seismic signals, as showed in Figure 2. This template waveforms are found using more basic detection methods, such as manual picking or STA/LTA. When best applied, a large variety of templates are used, reflecting the broad range of earthquake properties, e.g., magnitudes, source types, locations, etc. When a match is identified, it serves as an indicator of the presence of an earthquake. It relies on the computation of the normalised cross-correlation coefficient (NCC) between a set of chosen template waveforms and the continuous waveform recordings from seismic instruments [Gibbons and Ringdal, 2006, Mu et al., 2017, Yoon et al., 2015].

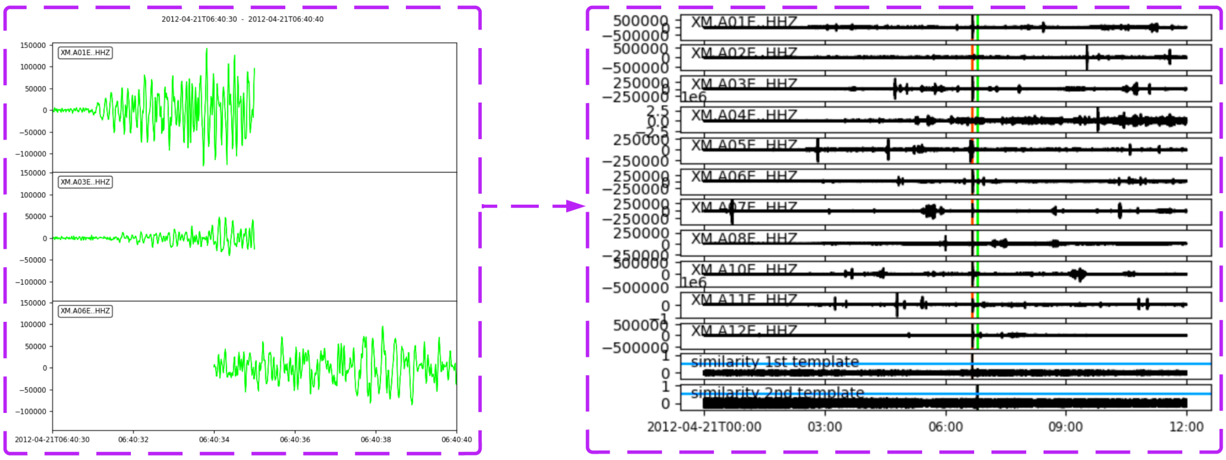

Figure 2: Multi-station detection of earthquake swarms employing multiple templates. The seismic dataset captures the initial 12 hours of a 2012 Aluto volcano earthquake event in the Main Ethiopian Rift. The seismic stream is high-pass filtered Z component. Extracted from the earthquake catalogue are the origin times and magnitudes of two primary earthquakes within this time-frame. Displayed in the left figure is a template waveform representing these earthquakes. Following this, cross-correlations are computed, aiding in the identification of additional similar earthquakes with cross-correlation at all stations.

Auto-correlation is a “many-to-many” search – referring to the process of systematically comparing a large number of waveforms with each other – for similar waveforms when the desired signal is not known in advance [Brown et al., 2008, Yoon et al., 2015]. The correlation between seismic waveforms from different monitoring stations is used to estimate the time delay and source location of seismic events, between stations. This can then be used to identify the onset of earthquakes detected by the stations in the network.

In addition to template matching and auto-correlation, migration methods can be employed for detecting and locating seismic events [Chambers et al., 2010, Langet et al., 2014]. Coalescence microseismic mapping (CMM) is one such method for both simultaneously detecting and locating seismicity [Drew et al., 2013, Winder et al., 2022]. In this approach a travel time look-up table (LUT) is established by forward modeling seismic arrival times from a defined grid to all receiver stations, using a input velocity model. This process is undertaken only once for a specific search volume and velocity model to accelerate the subsequent migration of signals. Travel time LUTs are generally created for both

P and S waves, and sometimes other reflected phases. Waveforms are then fed into the algorithm, with energy from the three components being migrated back to each grid point in the LUT to form onset functions for each time step. Thresholds in the magnitude of the onset function are then used to measure the phase arrival times, from which a preliminary location is computed. This location will be the point in the LUT with the highest coalescence value.

1.2 Challenges of conventional earthquake detection

Precisely establishing the arrival times of seismic P- and S-waves is a crucial stage in determining earthquake epicenters, magnitudes, and depths [Pardo et al., 2019]. However, manual picking is time consuming and many existing algorithms often fall short in terms of automation, distinguishing between multiple phases, and ensuring accuracy in performance [Cano et al., 2021].

The performance of STA/LTA can be hampered by various factors: low signal-to-noise ratio (SNR); emergent signals; complex waveforms; overlapping events; cultural noise; or sparse station spacing. Each of these results in a lower sensitivity for detecting earthquakes. This manifests as a fundamental limit in the size of earthquake that can be detected, as smaller or more distant events will naturally have a lower amplitude signal on a given seismic station. This limit is often termed the magnitude of completeness for a given station, array, or network of seismometers.

Template matching also has fundamental limitations since it relies on known waveform templates, making it less versatile for detecting sources that have not been seen before (e.g., from a different location or of a different faulting style). Functionally, TM will only find events with similar source characteristics to those already detected. Typically, templates are chosen from existing earthquake waveform data, which can naturally lead to biases in the resulting earthquake catalogue to similar event types.

The application of auto-correlation can sometimes have a similar limitations. Due to this method being computationally expensive, often smaller windows of data are used to compare to the larger datasets, biasing detections to a specific period. Auto-correlation can also generate many detections from noise sources that are recorded across arrays, meaning genuine earthquake detections have to be filtered out. This requires extra processing steps in the workflow. These effects naturally introduce limitations to the application of the above methods for earthquake detection for large continuous datasets [Yoon et al., 2015].

CMM is also somewhat computationally intensive, with travel-time look up tables needing to be created for a potentially large grid volume. This becomes a challenge for monitoring large spatial areas. It also requires relatively well constrained velocity models, also a challenge for large areas, to accurately pick and associate P and S times from the same events. As above, coherent noise sources would also be highlighted CCM detection routines, requiring further filtering and quality control to isolate. If the frequencies of the noise sources is not sufficiently distinct to the target earthquake signals, this can introduce many false detections into CMM-derived catalogues.

These challenges in conventional earthquake detection motivated the development of more computationally efficient methods that are less affected by noise and internal biases. Machine learning and deep learning methods employ many transformations and convolutions to input data, meaning they may circumvent some of these issues in a computationally efficient manner. Thus, these methods provide an avenue to develop novel detection methodologies. These approaches have the potential to enhance automation and accuracy, offering improved earthquake detection capabilities in a variety of complex and challenging scenarios. In the next section, AI and ML methods are introduced, and their application to earthquake detection and location is summarised.

2 Artificial intelligence, machine learning, and deep learning

In the recent years, the terms "artificial intelligence", "machine learning", and "deep learning" have become increasingly pervasive in discussions spanning various disciplines. These concepts have transcended their origins in computer science to impact fields as diverse as science and technology [Xu et al., 2021], education [Zawacki-Richter et al., 2019], agriculture [Eli-Chukwu, 2019], energy [Zahraee et al., 2016], engineering [Shukla et al., 2019], medicine [Briganti and Le Moine, 2020], healthcare [Haleem, 2023], manufacturing [Li et al., 2017], finance [Ahmed et al., 2022], marketing [Haleem et al., 2022], government [Ahn and Chen, 2020], and arts [Mazzone and Elgammal, 2019], and even our day-to-day life (e.g., smart assistants, health and fitness apps, predictive text, and auto-correct) [Xu et al., 2021]. While these terms are often used interchangeably, they do have slight differences in meaning that are important to understand. In this section, we aim to make these concepts more accessible, providing a basic understanding of what AI, ML, and DL entail and how they are impacting various areas. While we won't delve into intricate technical details, we will explore their differences, and the different types of ML and DL. We will also consider how these concepts explicitly relate to geophysics and seismology.

2.1 Artificial Intelligence (AI)

Artificial intelligence, often abbreviated as AI, refers to the development of computer systems and algorithms designed to perform tasks that would typically require human intelligence. These tasks encompass a wide range of activities, including problem-solving [Rattan et al., 2022], pattern recognition, language understanding, decision-making, and learning from experience. The ultimate goal of AI is to create machines and systems that can emulate human-like cognitive functions, enabling them to adapt, improve, and excel in a variety of domains [Xu et al., 2021]. There are multiple definitions for AI [Bini, 2018], but in computer science AI is the study of "intelligent agents", which are devices that "perceive their environment and take actions to maximise their chance of success at some goal" [Poole et al., 1998, Shinde and Shah, 2018].

AI draws its inspiration from human intelligence and cognitive processes. The inception of modern AI research can be attributed to John McCarthy, who introduced the term "artificial intelligence" during a lecture held at Dartmouth College in 1956 [Benko and Lányi, 2009, Flasiński and Flasiński, 2016, Bini, 2018, Wang, 2019]. This historical milestone marked the initiation of AI as a distinct field. Rooted in this history dating back to the mid-20th century, AI emerged as an academic discipline in the 1950s including the work of Alan Turing's "Computing Machinery and Intelligence", which raises the question of whether machines can "think" [Turing, 1950].

However, for over half a century AI remained a relatively obscure concept from both a scientific and practical standpoint [Haenlein and Kaplan, 2019]. AI experienced a resurgence in the 1980s when research institutions and universities developed AI systems capable of summarising expert knowledge into basic rules, aiding non-experts in decision-making processes [Xu et al., 2021]. This resurgence was further fueled in 2006 [Helm et al., 2020, Xu et al., 2021], marked by a surge in AI research activity and the emergence of deep learning algorithms of Hinton et al. [2006], Hinton and Salakhutdinov [2006]. Today, AI's renewed vitality can be attributed to factors such as the proliferation of so-called "big data" (data that contains greater variety, arriving in increasing volumes, and with more velocity), advancement in algorithms, advancement in computing powers, and others [Ergen, 2019].

Computing power, particularly graphical processing units (GPUs) [Martinez et al., 2019], and storage infrastructure play vital roles, providing the necessary computational resources to process and analyse vast data sets and store the models and data required for AI applications. Thus, the field of AI is advancing at unprecedented pace. As depicted in Figure 3 and 4, the substantial growth in publications, across all AI fields and by field of study respectively, serves as a testament to the accelerating pace of AI development.

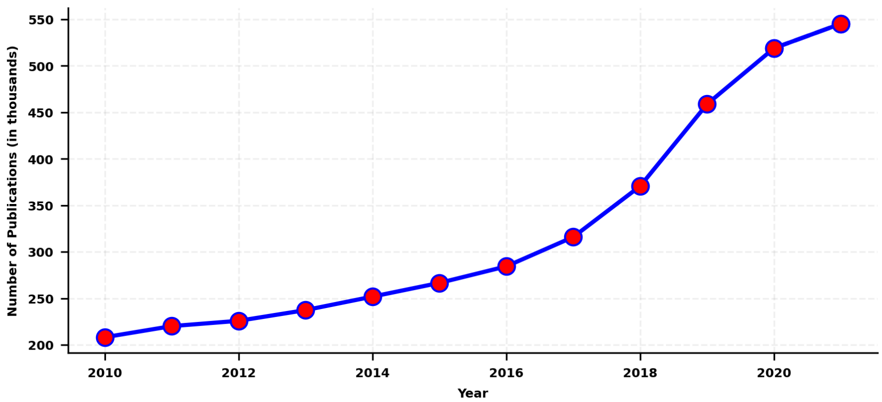

Figure 3: Evolution of AI publications from 2010 to 2021. This illustrates the increasing trend in the number of AI publications worldwide over the years based on data sourced from the Center for Security and Emerging Technology (September, 2023).

AI encompasses a broad framework that includes data collection and analysis, machine learning, natural language processing, computer vision, robotics, and ethical considerations, as shown in Figure 5. It aims to replicate human-like abilities in machines and with diverse applications.

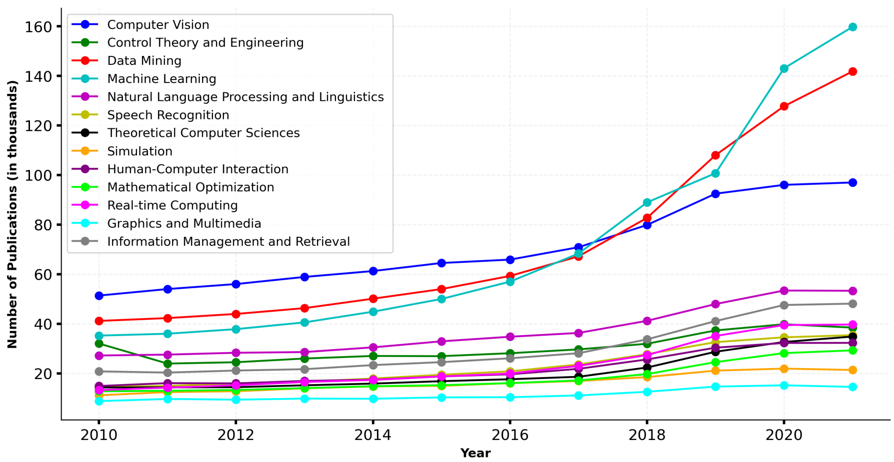

Figure 4: AI publications by field of study. from 2010 to 2021. This shows the number of publications across various field of study, highlighting the increasing trend in the field of machine learning compared to others based on data sourced from the Center for Security and Emerging Technology (September, 2023).

Figure 5 visually illustrates the distinctions between AI, ML and DL, showing the hierarchical relationship and roles within this field. The field of AI features a hierarchical relationship that encompasses ML and, within ML, DL. AI serves as the overarching framework that seeks to replicate human-like abilities in machines. ML represents a subset of AI, focusing on the development of algorithms that enables computers to learn from data. DL, in turn, is a specialised form of ML that employs deep neural networks with multiple layers, enabling it to model and solve complex problems. The hierarchy illustrates how AI encompasses a spectrum of technologies, with each level building upon the last, contributing to the evolution of "intelligent" systems.

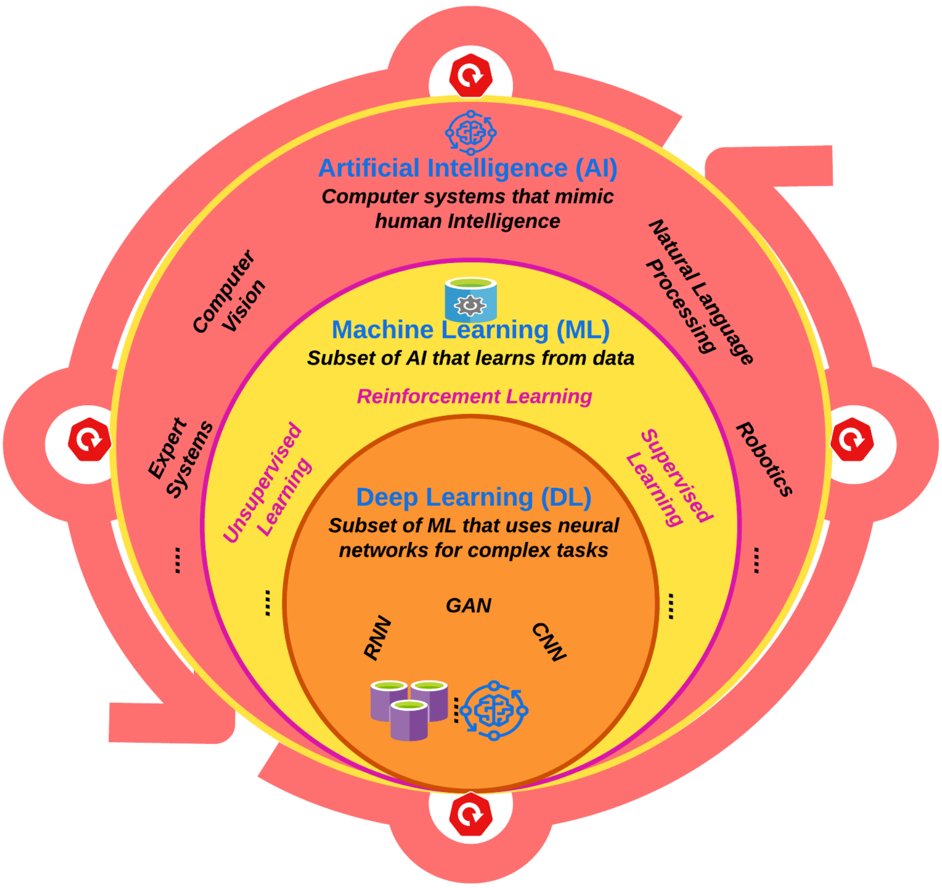

Figure 5: Hierarchical Relationship between AI, ML, and DL. The figure illustrates the hierarchical structure, with AI encompassing various fields, ML incorporating algorithms such as supervised, unsupervised and reinforcement learning, and DL comprising specific architectures like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Generative Adversarial Network (GAN).

2.2 Machine learning: a subset of AI

Machine learning, as a subset of AI [Wang et al., 2009], encompasses algorithms and statistical models employed by computer systems to accomplish specific tasks without requiring explicit algorithmic programming [Mahesh, 2020, Alzubi et al., 2018]. It explores to improve computer systems through experience [Alzubi et al., 2018, Ray, 2019] and understanding universal learning principles across computers, humans, and organisations [Jordan and Mitchell, 2015]. It empowers software applications to enhance their predictive accuracy through autonomous learning from data, eliminating the need for human intervention. In essence, it enables software applications to acquire knowledge independently by adhering to predefined instructions [Madakam et al., 2022, El Naqa and Murphy, 2015, Shalev-Shwartz and Ben-David, 2014].

The demand for ML functionality is growing rapidly [Mohr et al., 2018], and can be used in problems such as classification, regression, forecasting, anomaly detection, clustering, and dimensional reduction. The use of machine learning can be found in many fields such as robotics, computer vision, and language processing [Dash et al., 2020]. Figure 6 shows the machine learning cycle which is a sequence of stages that includes data collection, pre-processing, feature extraction, model training, evaluation, and deployment. It involves the process of transforming data into a model that can take predictions or decisions, and it often includes ongoing monitoring and refinement to maintain model performance.

Within machine learning, features are properties or characteristics of data that can be the identification or replication target of the algorithm. These are generally broken down into two types: numerical and categorical. Seismic data, as a measurement of amplitude over time, can be treated a relatively simple numerical feature. Other features that could be exploited in seismic data include the amplitude in given frequency bands, the ratio of amplitudes in time windows, and the correlation of time series over different ranges. Categorical features are generally abstracted to numerical features through some kind of encoding. Other examples of features in common machine learning methods include edges or size of objects in computer vision, and the shape and number of filled pixels in character recognition.

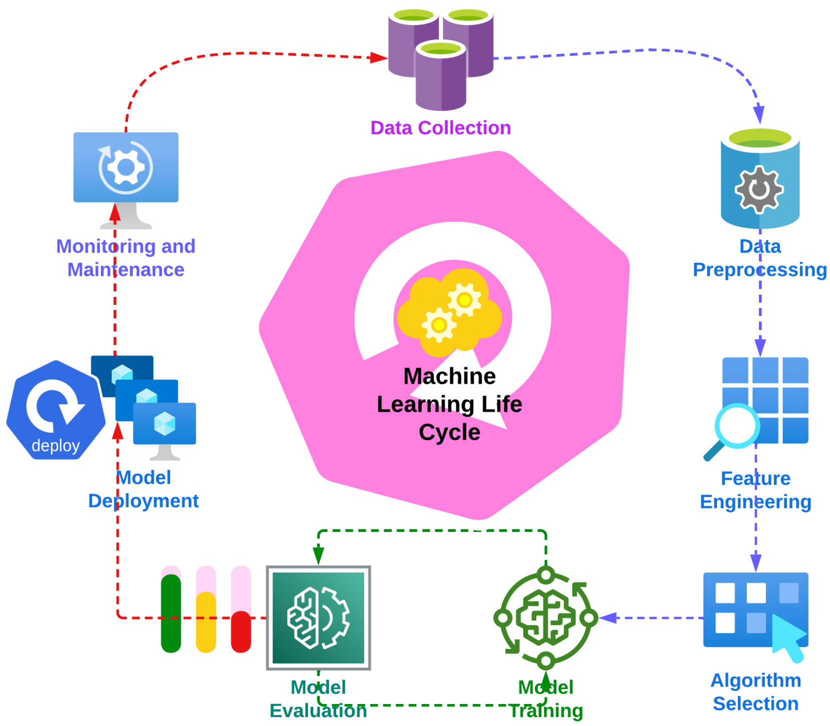

Figure 6: The iterative stages of machine learning, including data collection, processing, feature engineering, algorithm selection, model training, evaluation, deployment, and ongoing maintenance and monitoring, forming a cyclical process.

2.3 Definitions and types of ML algorithms

Machine learning relies on different algorithms to solve different problems [Mahesh, 2020]. These algorithms are categorised into a taxonomy based on the desired outcome of the algorithm [Osisanwo et al., 2017]. Three primary categories of ML algorithms exist: supervised; unsupervised; and reinforcement [Pandey et al., 2019, Singh et al., 2016, Singh and Singh, 2019, Praveena and Jaiganesh, 2017]. Additionally, there is a fourth category known as semi-supervised, which emerges from the fusion of supervised and unsupervised approaches.

Notable algorithms within the ML domain include: transduction; learning to learn; evolutionary learning; ensemble learning; instance-based learning; dimension reduction algorithms; and hybrid learning [Nasteski, 2017, Ayodele, 2010, Alzubi et al., 2018]. Although, there is no immediate answer to which machine learning algorithm is right for a specific task, several factors can help refine the selection. These factors include the interpretation of the algorithm, number of data points and features, data format, linearity of data, training time, prediction time, and memory limitations.

2.3.1 Supervised learning

Supervised learning is one of the fundamental methods in machine learning. In this type of learning, the algorithm is provided with a labelled data-set, which consists of input data paired with corresponding target labels or outcomes. The goal of supervised learning is to learn a mapping function that can predict the target labels accurately of unseen data [Muhammad and Yan, 2015].

This learning category can perform mainly two types of tasks: classification and regression [El Mrabet et al., 2021]. Supervised methods are used in various applications such as marketing, finance, manufacturing, and stock market prediction [Praveena and Jaiganesh, 2017]. The most common supervised learning techniques are decision tree, k-nearest-neighbor, support vector machines, and naive Bayes algorithms [Mohamed, 2017, Alloghani et al., 2020]. The schematic workflow of supervised machine learning algorithms is given in Figure 7.

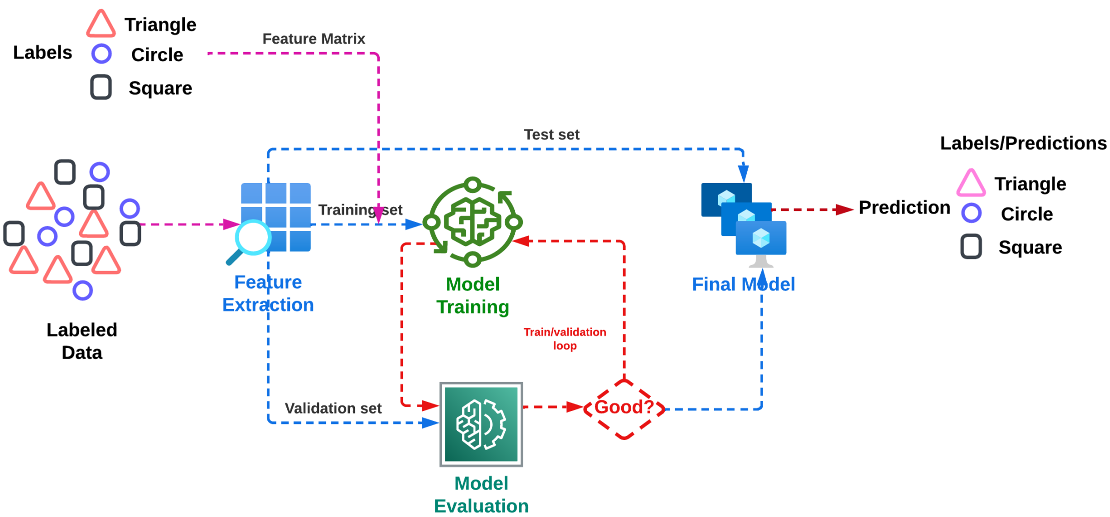

Figure 7: The stages of supervised machine learning as shown, involving the input of labelled data and feature extraction, followed by training, testing, and validation. The process includes a train-validation loop (red-dashed line), leading to the development of a final model and concluding with prediction. The labelled data is combined into a matrix structure, with features (i.e., attributes or properties) converted into numerical values, in order to use it during the feature extraction step. This is known as the feature matrix.

2.3.2 Unsupervised learning

In supervised learning, the algorithm is trained on a data set consisting of instances paired with their corresponding known labels or outputs. In contrast, unsupervised learning deals with data sets where instances lack predefined labels and are therefore unlabelled [Kotsiantis et al., 2007, Bhavsar and Ganatra, 2012, Palacio-Niño and Berzal, 2019].

This type of machine learning algorithm arguably represents AI in its purest application, removed from biases that can be introduced during the labelling process, and offers promising avenues to learn from an abundance of unlabelled visual data [Chen et al., 2022]. Examples of unsupervised learning algorithms include apriori algorithms, ECLAT, frequent pattern growth, k-means clustering, and principal component analysis [Naeem et al., 2023]. K-means, hierarchical clustering, and principal component analysis emerged as the most commonly used unsupervised techniques [Alloghani et al., 2020]. Unsupervised machine learning finds applications in diverse fields, including customer segmentation, anomaly detection [Omar et al., 2013], recommendation systems, image analysis [Raza and Singh, 2021], bio-informatics [Greene et al., 2014], and data visualisation. It extracts patterns from unlabelled data, enabling data-driven decision-making. The workflow of unsupervised machine learning algorithms is shown in Figure 8.

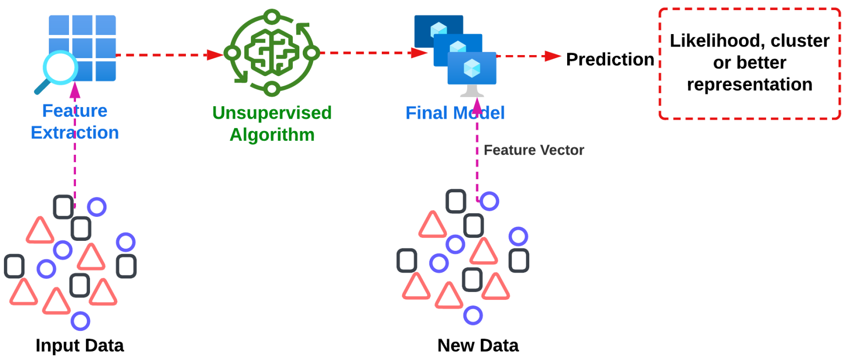

Figure 8: The stages of unsupervised machine learning, starting with input data and feature extraction, selection of unsupervised algorithm and the development of a final model. The final model finally make predication's on new data. The model can then be validated against a subset of the training data which is held back from the input data, or some other benchmarking algorithm.

2.3.3 Reinforcement learning

Reinforcement learning is a subset of machine learning that focuses on training intelligent "agents" to make sequential decisions in an environment in order to maximise a cumulative "reward" [Sutton and Barto, 2018, Thrun and Schwartz, 1994]. Unlike supervised learning, where the model is provided with labelled data, and unsupervised learning, which deals with finding patterns in unlabelled data, reinforcement learning relies on a feedback loop.

"Agents" learn by interacting with their environment, receiving rewards or penalties based on their actions, and adjusting their "behaviour" over time through trial and error [Kaelbling et al., 1996, Sutton, 1992, Barto, 1997, Glorennec, 2000, Qiang and Zhongli, 2011]. The agent and environment are the basic components of reinforcement learning [Ding et al., 2020]. The environment is an algorithm that the agent, a weighted set of variables, can interact with, where the reward can reinforce inherent behaviours in the model towards a specific outcome. This is shown schematically in Figure 9.

This paradigm has found applications in various fields, from robotics [Kober et al., 2013] and game-playing AI to autonomous driving [Shalev-Shwartz et al., 2016] and recommendation systems, showcasing its versatility and potential to tackle complex decision-making tasks. The use of deep learning is enabling the scaling up of reinforcement learning methods (in deep reinforcement learning, DRL) [Arulkumaran et al., 2017].

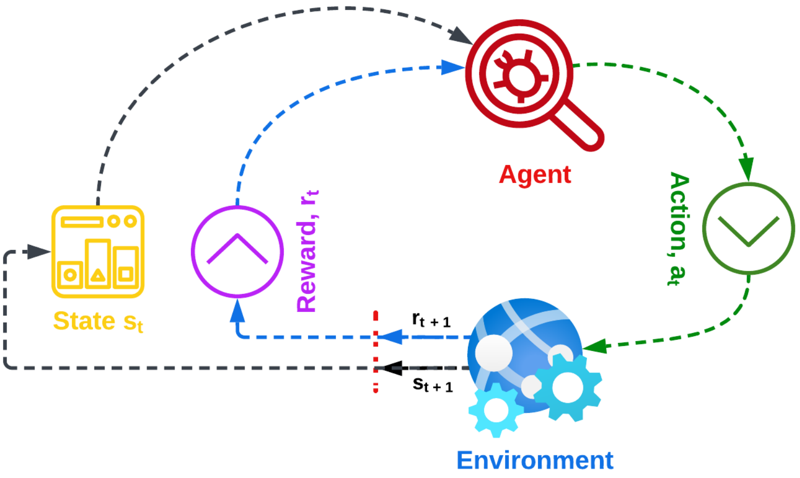

Figure 9: Reinforcement learning framework: a_t is the action taken by the "agent" at time t, s_t is the state of the environment at time t, and r_{t+1} and s_{t+1} stands for the reward and state at time $t+1$ respectively.

2.4 Machine learning applications

Machine learning has found many real-world applications across a wide range of fields [Sarker, 2021b, Shinde and Shah, 2018]. In healthcare [Triantafyllidis and Tsanas, 2019], it assists in disease diagnosis and personalised treatment plans by analysing medical records and imaging data. Financial institutions use machine learning for fraud detection [Thennakoon et al., 2019] and risk assessment, enhancing security and optimising investments. In the automotive sector, self-driving cars rely on machine learning algorithms to navigate safely and efficiently [Devi et al., 2020]. E-commerce platforms employ recommendation systems that use machine learning to suggest products, enhancing user experience and driving sales [Pallathadka et al., 2023]. Natural language processing [Otter et al., 2020] enables virtual assistants to understand and respond to human language, while image recognition powers facial recognition systems for security and accessibility. Climate scientists use machine learning to model and predict climate patterns, aiding in climate change mitigation efforts [Bochenek and Ustrnul, 2022]. In essence, machine learning has permeated many industries [Larrañaga et al., 2018].

2.5 Deep learning: a sub-field of ML

Deep learning is a subset of machine learning based on artificial neural networks [Janiesch et al., 2021, Li, 2017, Guo et al., 2016, Cai et al., 2020]. The growth in data volume and computing capabilities has led to increased interest in and the application of neural networks with more intricate architectures across various domains [Hao et al., 2016]. In recent years, deep learning has emerged as the dominant computational method within machine learning across a wide range of tasks [Alzubaidi et al., 2021]. It can often exceed human-level performance [Taye, 2023].

Deep neural networks with several layers have recently become a common focus of ML research due to their performance gains [Karhunen et al., 2015, Ongsulee, 2017]. Deep learning techniques include convolution neural networks, recurrent neural networks, autoencoders, deep-belief networks, recursive neural networks, and direct deep reinforcement learning [Shinde and Shah, 2018, Sarker, 2021a]. Due to their frequent use in earthquake detection methods, some of these deep learning approaches are introduced in the following sections.

2.5.1 Deep learning: definitions and significance

Deep learning is a sub-field of machine learning that focuses on training artificial neural networks to automatically learn and extract intricate patterns and representations from large and complex datasets [Vargas et al., 2017]. Inspired by the human brain [Shrestha and Mahmood, 2019], deep learning models consist of multiple layers of interconnected nodes that process and transform data hierarchically [Guo et al., 2016].

These models have demonstrated remarkable success in various domains, including image and speech recognition [Rusk, 2016], natural language processing [Du et al., 2016], computer vision [Voulodimos et al., 2018], and reinforcement learning [Mahmud et al., 2018]. Their ability to automatically discover and represent intricate features makes deep learning a powerful tool for tasks like image classification [Tamuly et al., 2020], language translation, and autonomous decision-making.

The biological neural network is a highly organised system that efficiently processes information from various senses using interconnected neurons. Artificial neural networks in machine learning replicate this structure to learn and recognise patterns in data [Suk, 2017]. Neural networks here refer to the artificial neural networks rather than biological ones and the computational units are connected to one another through weights, factors applied to the node that are changed during the training process [Aggarwal et al., 2018].

These networks consist of interconnected nodes, or artificial neurons, organised into layers that process and transform data. Each neuron is a mathematical processing unit and with all other neurons it learns the relationship between the input features and output [Georgevici and Terblanche, 2019]. The adaptability and ability to capture intricate features make neural networks a cornerstone of modern AI.

Single-layer networks use a modified linear function (a "perceptron"), while multi-layer networks arrange neurons into feed-forward layers, including input, hidden, and output layers [Choi et al., 2020]. Feed-forward layers are defined by one-way flow of information between their layers. This means that information moves strictly in a single direction from input nodes through hidden nodes and finally to the output nodes.

2.5.2 Deep learning vs. "traditional" ML

Machine learning traditionally relied on manual feature engineering, where training and architectures of layers were tuned to identify or replicate particular properties of the input data. Deep learning uses neural networks (NNs) to automatically extract complex representations from raw data. Thus, DL can perform better in so-called unstructured data tasks [Janiesch et al., 2021, Alom et al., 2019]. Deep learning is successful in areas like image recognition and natural language processing [Pouyanfar et al., 2018], but traditional machine learning is still relevant for structured data and interpretable models where transparency is vital. Transparency here means that the inner workings of the model are more understandable and explainable, making it easier for humans to comprehend how the model arrives at its decisions. Generally, deep learning techniques are now outperforming more traditional machine learning methods [Mathew et al., 2021, Chauhan and Singh, 2018].

Deep learning and traditional machine learning differ significantly. Deep learning, especially deep neural networks, excel with large datasets, automatically learning complex patterns. Traditional machine learning methods are generally more bound by the limits of the training data, with significantly reduced performance outside of these limits [Mathew et al., 2021, Khamparia and Singh, 2019].

Deep learning models often require extensive computational resources (and thus time) for training, particularly with complex NN architectures and large datasets [Sarker, 2021a]. In contrast, traditional machine learning models can be trained more quickly since they don't require the intensive optimisation process of deep learning. DL can often be economically expensive to conduct due to the need for powerful computer hardware, like graphics processing units (GPUs) or tensor processing unit (TPUs) [Sarker, 2021a].

Most importantly, deep learning models are often seen as "black boxes" [Gupta et al., 2023]. Their ability to distort the input data and find patterns through the setting of a very large number of parameters make it highly challenging to interpret their decisions. In contrast, traditional machine learning models, with simpler algorithms and more interpretable features, are preferred for scenarios where explainability of the model is crucial.

2.6 Deep learning architectures for earthquake detection

Discriminative deep learning aims to learn the boundary or decision boundary that separates different classes or categories data. These models are primarily used for classification tasks, where the goal is to map input data to a specific class label. Examples include the multi-layer perceptron (MLP), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) [Sarker, 2021a]. Discriminative models focus on modelling the conditional probability of the class given the input data, making them well-suited for tasks requiring precise classification and prediction [Deng, 2012].

2.6.1 Multi-layer perceptron

An MLP is a fundamental architecture in deep learning, composed of multiple interconnected layers of artificial neurons. MLPs consist of an input layer, one or more "hidden" layers, and an output layer [Delashmit et al., 2005, Singh and Banerjee, 2019]. This is shown in Figure 10. Each neuron in a layer that processes input data, applies a non-linear activation function, and passes its output to neurons in the next layer [Gardner and Dorling, 1998]. Through a process called backpropagation, MLPs learn to adjust the weighs of each node during training to make accurate predictions or classifications [Popescu et al., 2009]. Backpropagation adjusts the model weights by minimising the difference between the predicted and actual outputs, facilitating effective learning [Rumelhart et al., 1986].

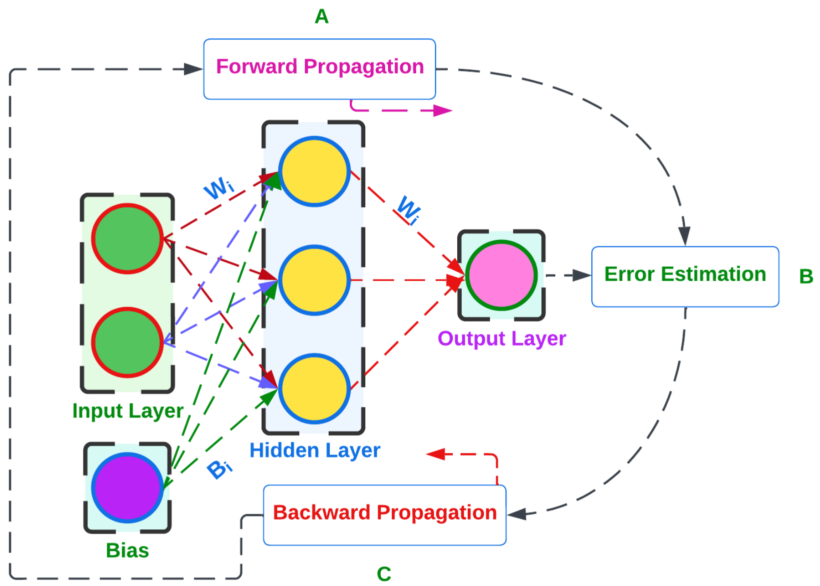

Figure 10: Multi-Layer perceptron (MLP) structure. The figure illustrates the architecture of a multi-layer perceptron, comprising an input layer, bias, hidden layers, and an output layer. The connections between layers, represented by weights, contribute to the neural network's learning and predictive capabilities. Training involves adjusting weights (W_i) and biases (B_i) to minimise error by making corrections between input-output pairs. Backpropagation, through forward (A) and backward (C) passes, compute gradients, guiding parameter adjustments towards minimising errors (B), often employing optimisation algorithms (black-dashed loop)

2.6.2 Convolutional neural networks

CNNs have become one of the most representative architectures in DL [Li et al., 2021, Gu et al., 2018, Albawi et al., 2017]. They are designed for processing and analysing visual data, such as images and videos. CNNs are inspired by the human visual system, featuring layers of neurons that apply convolutional operations to input data. These operations involve sliding small filters or kernels over the input, capturing local patterns, and gradually extracting higher-level features as they progress through the information in the image or data series [Yamashita et al., 2018]. They are comprised of the convolutional layers, pooling layers, and the fully-connected layers [O'Shea and Nash, 2015]. This is shown in Figure 11.

The classification step shown in Figure 11 functions like a tradiational machine learning algorithm. The flattening stage is conducted to convert the 2-dimensional convolutional matrices into a feature vector, which the fully-connected layers and activation functions can act upon.

CNNs have undergone various developments and spawned several important variants that have improved their performance in different applications. Some notable CNN variants include, VGGNet [Simonyan and Zisserman, 2014], AlexNet [Krizhevsky et al., 2012], GoogLeNet (Inception) [Szegedy et al., 2015], ResNet (Residual Network) [He et al., 2016], DenseNet [Huang et al., 2017], MobileNet [Howard et al., 2017], and Xception [Chollet, 2017]. These CNN variants have generally advanced the field of computer vision, including image recognition, object detection, and image segmentation. New CNN architectures and adaptations are developed to tackle different tasks, though many of the above methods can be used in more general ML applications.

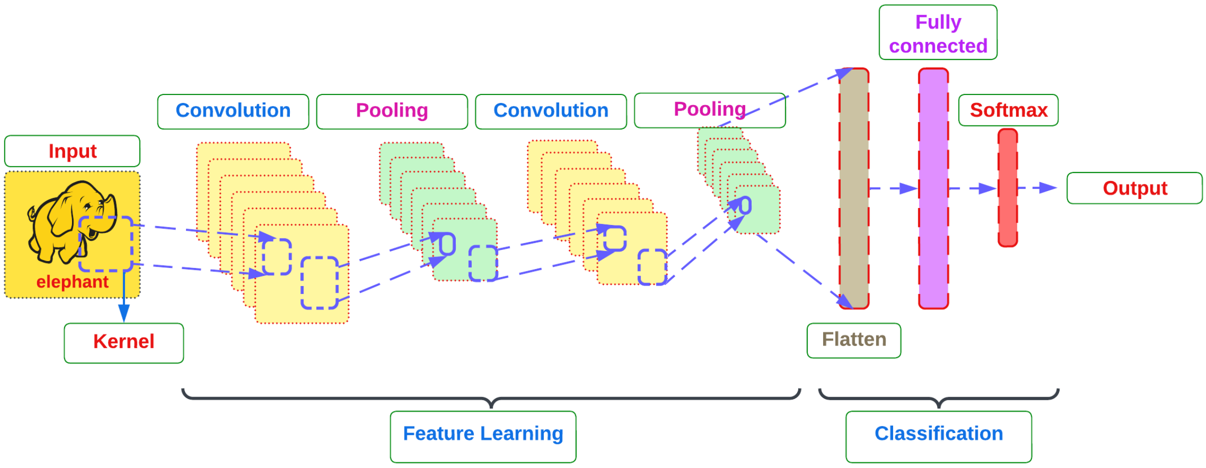

Figure 11: Convolutional Neural Network (CNN). The figure illustrates the sequential components of CNN, featuring input with a kernel, convolution layers, pooling layers for feature learning, followed by flattening, fully connected layers, and a softmax activation function layer for classification, ultimately leading to the network's output.

2.6.3 Recurrent neural networks

RNNs have been an important focus of research and development during the 1990's [Medsker and Jain, 2001]. RNNs are a class of artificial neural networks designed for processing sequences of data [Salehinejad et al., 2017, Sutskever et al., 2011, Lipton et al., 2015]. They are particularly well-suited for tasks like natural language processing, speech recognition [Graves, 2013, Delashmit et al., 2005], and time series analysis [Yu et al., 2019], hence their use in earthquake detection.

RNNs are characterised by their ability to maintain hidden states, which allows them to capture information from previous time steps and use it to influence their current predictions. However, traditional RNNs suffer from the vanishing gradient problem, which limits their ability to capture long-range dependencies in sequences. Vanishing gradients in RNN occur when, during backpropagation from the output to the input layer, the gradient diminishes exponentially, causing the weights of earlier layers to remain nearly unchanged. This occurs in circumstances where the chosen activation functions lead to outputs with very small changes with respect to the change in weights during the backpropagation step.

The vanishing gradient problem hinders the learning of long-term dependencies, preventing gradient descent from converging to the optimum. To address this issue, several variants have been developed, such as long short-term memory (LSTM) networks, bidirectional LSTMs (BiLSTMs), and gated recurrent units (GRUs) [Sutskever, 2013].

LSTM networks use a more complex gating (refers to a mechanism that regulates the flow of information within the network) mechanism to better preserve and update information over time. BiLSTMs process sequences in both forward and backward directions, enhancing their understanding of context [Schuster and Paliwal, 1997]. GRUs offer a simplified gating mechanism compared to LSTMs [Dey and Salem, 2017], which can make them more computationally efficient [Fu et al., 2016] while still addressing the vanishing gradient problem. These variants have become essential tools in deep learning, enabling the modeling of sequential data with greater accuracy and efficiency.

2.6.4 Generative networks

Generative deep learning focuses on generating new data that resembles the training data. These models have applications in image generation, text-to-image synthesis, and data augmentation.

Generative adversarial networks (GANs) are class of deep learning models introduced by [Goodfellow et al., 2014]. GANs are designed to address the challenge of generating realistic and high-quality data, whether it be images, text, audio, or other forms of content. They consist of two neural networks, a generator, and a discriminator [You et al., 2022], engaged in a competitive, adversarial process [Goodfellow et al., 2014, Aggarwal et al., 2021, Gonog and Zhou, 2019]. The generator aims to create data that is indistinguishable from real data, while the discriminator attempts to differentiate between genuine and generated data as it is shown in Figure 12. Through iterative training, GANs learn to produce increasingly authentic output, leading to impressive results in image synthesis, style transfer, super-resolution, and more. GANs have recently achieved impressive results in real world applications [Bau et al., 2018, Jabbar et al., 2021]. They have had a profound impact on the fields of computer vision, natural language processing [Durgadevi et al., 2021], and generative art, and they continue to drive innovation in the realm of AI and creative content generation.

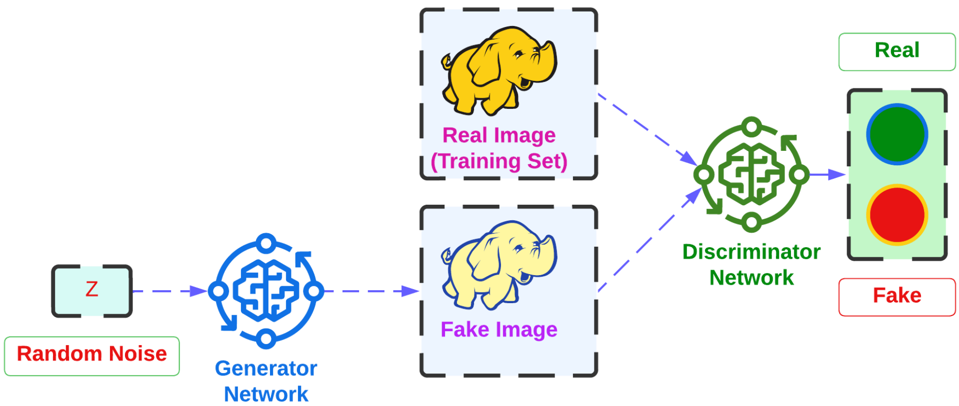

Figure 12: The generative adversarial network (GAN) framework, consisting of random noise input, a (in this case pre-trained) generator network creating fake images, and a discriminator network distinguishing between real images from a training set and generated fake images. The discriminator provides an output indicating whether the input is real or fake. The training of GANs would lead to the "fake" image shown here becoming more similar to the "real" image through each iteration.

AE are a class of neural networks designed for unsupervised learning and data compression [Sewak et al., 2020]. They consist of an encoder, which compresses input data into a lower dimensional representation (a "code"), and a decoder, which attempts to reconstruct the original data from this compressed representation [Bank et al., 2023, Pinaya et al., 2020]. This is shown in Figure 13.

Autoencoders have a wide range of applications, including dimensionality reduction, denoising, and anomaly detection. This makes them suited to filtering tasks in seismology. Variants of autoencoders have been developed to address specific challenges and extend their capabilities. Some notable variants include variational, sparse, denoising, contractive, and stacked autoencoders [Bank et al., 2023, Burda et al., 2015].

Autoencoders and their variants are powerful tools for unsupervised learning and data transformation. They find applications in a broad range of domains. including computer vision, natural language processing, and anomaly detection, by capturing meaningful representations of complex data.

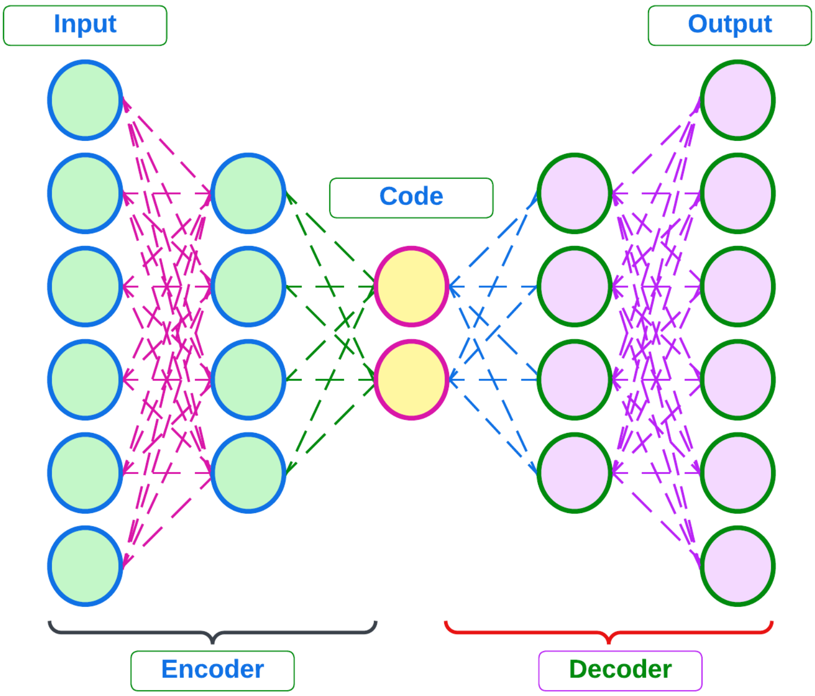

Figure 13: Autoencoder. The structure of an autoencoder, including the input, encoder, code, decoder, adn output layers. The autoencoder is designed for unsupervised learning, aiming to encode and then reconstruct the input data, effectively capturing essential features for representation.

2.6.5 Hybrid and other architectures

Hybrid deep learning approaches combine elements of both discriminative and generative models to leverage their respective strengths. These models aim to capture the joint distribution of data and latent variables, making them versatile for a wide range of applications. Hybrid models are often used when there is a need for both generative and discriminative precision in a single model [Sarker, 2021a].

2.6.6 Deep transfer learning

Transfer learning (TL) is a machine learning technique that leverages knowledge gained from one task to improve the performance of a related task [Pan and Yang, 2009, Weiss et al., 2016, Zhuang et al., 2020]. It involves taking a pre-trained model, typically trained on a large dataset for a specific task, and fine-tuning it on a new task with a smaller training dataset [Yang et al., 2020]. This approach allows the model to inherit useful features and representations learned during the initial training, which can significantly speed up training, enhance acurracy (e.g., see Hattula et al. [2023] and the model's performance on the new task [Niu et al., 2020].

Deep transfer learning (DTL) is an advanced extension of transfer learning, primarily applied in the context of deep neural networks and it reduces the reliance on extensive labelled data and training costs [Iman et al., 2023]. In DTL, not only the final layers of the pre-trained model are fine-tuned but also intermediate layers are adapted to the new task. This technique is particularly valuable when dealing with complex tasks and large datasets, as it enables the transfer of deeper and more abstract representations from the source task to the target task. DTL has found widespread applications in various domains, such as computer vision, natural language processing, and reinforcement learning [Taylor and Stone, 2009, Zhu et al., 2023c].

2.6.7 Deep reinforcement learning

Deep reinforcement learning (DRL) combines DL with reinforcement learning principles enabling machine learning models to make sequential decisions in complex environments [François-Lavet et al., 2018]. DRL agents learn through trial and error, just like traditional reinforcement learning, but they use deep neural networks to approximate complex, high-dimensional, state-action mappings (the relationships between the current state of the system and the actions that the reinforcement learning can take in that particular state) [Li, 2023]. This allows DRL systems to handle a wide range of tasks, from game playing [Arulkumaran et al., 2017] and robotic control [Zhao et al., 2020] to autonomous vehicles [Kiran et al., 2021] and recommendation systems.

2.6.8 Transformers

Transformers are a class of deep learning models that have had a profound impact on natural language processing [Wolf et al., 2020] and a wide range of other machine learning tasks. Introduced in Vaswani et al. [2017], transformers have revolutionised the field by using self-attention mechanisms to process sequences of data, such as words in a sentence or pixels in an image, in parallel rather than sequentially. This parallelisation greatly accelerates training and improves performance on tasks like machine translation, text summarisation, sentiment analysis, and more. The transformer architecture has since evolved into various forms, such as BERT [Devlin et al., 2018], and GPT [Brown et al., 2020, Radford et al., 2018], each optimised for specific tasks.

3 Machine learning in seismology

The rise of AI in geophysics is a rapidly growing phenomenon, transforming the way geophysicists conduct research and analysis. Neural networks are gaining popularity in geophysics [Van der Baan and Jutten, 2000]. AI technology is being applied in seismic interpretation [Wang et al., 2018, Silva et al., 2019], speeding up the analysis of seismic data and providing valuable insights for oil and gas exploration [Sircar et al., 2021, Tariq et al., 2021] and mineral exploration [Shirmard et al., 2022]. AI also plays a crucial role in hazard assessment [Gitis and Derendyaev, 2019, Harirchian et al., 2020], predicting natural disasters, aiding disaster management efforts and others. In the field of climate science [Rolnick et al., 2022, Monteleoni et al., 2013], AI models analyse complex climate data to make accurate predictions about future climate trends. Challenges include data quality, data availability and others. Despite these challenges, the potential benefits of AI in geophysics are substantial, making it an integral part of the field's future.

Seismology studies the propagation of elastic waves to study their sources, including earthquakes and explosions, and the structure of Earth. Seismological studies record data across a broad range of frequencies to study structures of different length scales. In modern seismology, large numbers of sensors are used to imporve the resolution of many of the underlying imaging methods. Seismology is a data-rich field, which makes it an ideal domain for ML [Kong et al., 2019]. This large increase in the volume of data used in seismology has made it natural to employ the use of ML. These techniques can help quickly analyse the large quantities of data gathered from sensors.

3.1 Data and preprocessing

ML and DL algorithms, although diverse in their implementation, tend follow a basic workflow that includes data collection and preprocessing [Kong et al., 2019]. Seismic data typically collected from a network of sensors or seismometers, is often large and complex. To render this data suitable for ML analysis or training, a series of preprocessing steps is necessary. These may include data cleaning to eliminate noise, re-sampling to ensure uniform time intervals, and filtering to enhance the signal-to-noise ratio.

Significant data preprocessing does add an additional challenge when applied to earthquake detection. It is likely that for many applications where the performance of traditional detection methods is deemed satisfactory, many of the ML methods described will not be used. The additional information gained may not be considered sufficient to justify the additional work required to properly tune, apply, and interrogate the ML techniques. This is a judgement to be made on a dataset-by-dateset basis.

Feature engineering is a critical step when using ML on seismic data, where domain-specific attributes such as spectral characteristics or time-frequency representations are calculated to capture relevant information. Moreover, data augmentation techniques can be applied to expand the seismic dataset, particularly when dealing with limited seismic event occurrences.

3.2 ML and DL applications in geophysics

ML techniques are becoming increasingly widespread in seismology, with various applications [Kong et al., 2019, Bergen et al., 2019]. DL particularly has attracted increasing attention in the geophysical community, resulting in many opportunities and challenges, as described below.

ML and DL can significantly improve our capability for seismic data processing [Mousavi and Beroza, 2023] and have been used in various applications. Some of the applications of ML and DL in seismology include, seismic lithology prediction [Zhang et al., 2018], seismic data inversion [Li et al., 2019, Zheng et al., 2019, Zhang et al., 2021], phase picking [Wang et al., 2019, Woollam et al., 2022, Chai et al., 2020, Lapins et al., 2021, Zhu and Beroza, 2019, Mousavi et al., 2020, Soto and Schurr, 2021, Ross et al., 2018a, Bornstein et al., 2023], phase detection [Ross et al., 2018b, Reynen and Audet, 2017, Yoon et al., 2015, Chen and Li, 2022], seismic phase association [Ross et al., 2019, McBrearty and Beroza, 2023], event classification [Trani et al., 2022], fast simulation of seismic waves in complex media [Moseley et al., 2020], full-waveform inversion [Liu et al., 2021], earthquake early warning [Li et al., 2018], earthquake monitoring [Zhu et al., 2023b], tomography [Bianco et al., 2019], earthquake forecasting [Beroza et al., 2021], microseismic monitoring [Anikiev et al., 2023], distributed acoustic sensing (DAS) [Shiloh et al., 2019, Hernández et al., 2022, Stork et al., 2020, Zhu et al., 2023a, Van den Ende and Ampuero, 2021, Batson and Royer, 2019], and carbon capture and storage (CCS) [Yao et al., 2023, Wen et al., 2021, Menad et al., 2019, Kaur et al., 2023].

3.2.1 Phase picking and earthquake detection

Phase picking plays a fundamental role in seismic analysis as it serves as the cornerstone for all further processes [Zhu and Beroza, 2019, Guo et al., 2020], including tomography, source characterisation, and others. The precise identification of seismic phases is the foundation upon which seismic studies are built.

Phase picking and earthquake detection have recently been the focus of ML research [Jiao and Alavi, 2020]. ML and DL have improved performance in the field of phase picking and earthquake detection by automating seismic event identification. The main models used are explained in the next section.

In phase picking, ML algorithms can analyse seismic waveforms to identify the arrival times of various phases, such as P- and S-waves. DL, with its neural networks and hierarchical feature extraction, has proven particularly effective in discerning subtle patterns in seismic data. These techniques can rapidly process and classify seismic signals, distinguishing genuine earthquake events from background noise. As the volume of seismic data continues to grow, the synergy between ML, DL and seismology holds great promise for advancing our understanding of the subsurface and response to earthquakes.

3.3 Deep learning architectures for detection

Deep learning architectures and models have brought a new dimension to earthquake detection, particularly phase picking. Several models have been developed to learn complex patterns and features within specific databases of seismic waveforms, allowing for accurate and efficient identification of arrivals. DL models have also been adapted to improve performance over time, or use in an alternative setting. As models are "exposed" to more seismic data, this continuously refines and optimises phase picking algorithms to become more versatile, ultimately contributing to more precise seismic event detection and analysis over a broader geographical range.

The most often used deep learning architectures in seismology are CNNs and RNNs [Mousavi and Beroza, 2022]. Both have been applied to automatic phase picking, with CNNs being used for their ability to capture spatial features in seismic waveforms, and RNNs handling temporal dependencies. Furthermore, newer architectures like transformer based models [Mousavi et al., 2020], which excel in sequence-to-sequence tasks, are showing promise for earthquake detection and source parameter estimation by effectively handling time-series data.

U-Net, a specialised CNN architecture, excels at precisely delineating object boundaries and capturing fine-grained details within images, making it a powerful tool for different applications where accurate segmentation is crucial. This architecture has had a significant impact on the field of image segmentation and continues to be widely used for various pixel-wise classification tasks. When examining seismic data in the form of images – a method that has been pplied to, for example, fibre optic distributed acoustic sensing data – this particular architecture has been shown to be quite effective [Stork et al., 2020].

3.3.1 Transfer Learning

Transfer learning has also been applied in the domain of phase picking. This approach leverages the knowledge gained from broader seismic data and generalises it to specialised context of other datasets. Lapins et al. [2021] apply this approach to volcano seismology. They adapted these pre-trained models, improving the accuracy and efficiency of phase picking for volcanic earthquakes, which often present unique challenges due to their distinct seismic signatures. This method not only accelerates the development of robust models but also reduces the need for extensive labelled data, and computing resources.

3.4 Current ML detection algorithms

The most prominent of the ML-based phase detection algorithms are EQTransformer [Mousavi et al., 2020], GPD [Ross et al., 2018b], and PhaseNet [Zhu and Beroza, 2019]. Variants and other models include U-GPD [Lapins et al., 2021], CRED [Mousavi et al., 2019], DPP [Soto and Schurr, 2021], and BasicPhaseAE [Woollam et al., 2019]. These algorithms, otherwise known as models, are trained on various seismic data and tectonic environments, and have advanced the automation and accuracy of phase picking in seismic waveform data. Detailed information about their architectures and performance can be found in the cited articles, and brief summaries of the most prominent models are given below. Table 1 also provides a condensed overview of each model's specific features.

Some of the models summarised here deal with earthquake detection, some phases picking, and some both. This distinction is generally a matter of accuracy. Algorithms which just do detection will give areas of the time series where an event signal is likely to have been found. However, phase picking models will also give the incident time, and often associate, P-, S-, and sometimes other phases. These picking models will have been trained on pick time data for specific phases, manually labelled from large earthquake databases.

EQTransformer is a deep learning model designed for simultaneous earthquake detection and phase identification. This model is adept at recognising earthquake signals and accurately determining the initial P and S phases in single-station data recorded at local epicentral distances (<300 km), leveraging an attention mechanism. The neural network follows a multi-task structure, featuring a very-deep encoder and three distinct decoders that incorporate a range of architectural elements, including 1D convolutions, bi-directional and uni-directional long-short-term memories (LSTM), residual connections, feed-forward layers, transformer and self-attentive layers.

For training, EQTransformer uses the STanford EArthquake Datasets (STEAD), a comprehensive global dataset containing manually labelled earthquake and non-earthquake signals. The dataset is partitioned randomly into training (85 percent), validation (5 percent) and test (10 percent) sets. The waveform are 1 minute duration, sampled at 100 Hz, and causally band-pass filtered within the range of 1 to 45 Hz [Mousavi et al., 2020].

GPD, a convolutional neural network (ConvNet), has been trained on extensive hand-labelled datasets from the Southern California Seismic Network to identify seismic body-wave phases. This approach, known as generalised phase detection (GPD), excels in reliably detecting P and S waves across a broad range of earthquake magnitudes, eliminating the need for explicit waveform templates. The preprocessing steps involved de-trending and high-pass filtering data above 2 Hz to eliminate microseismic noise. The record has a fixed duration of 4 seconds (400 samples) [Ross et al., 2018b].

PhaseNet is a model designed for determining the arrival times of P and S waves in seismic waveforms. Using three-component seismic waveforms as input, PhaseNet employs a modified U-Net architecture tailored for 1-D time-series data. The model produces probability distributions for P and S arrivals, and background noise as its output. The training process involves a substantial dataset with over 700,000 waveform samples extracted from more than 30 years of earthquake recordings, labelled by analysts at the Northern California Earthquake Data Center. The input and sequence consist of 3001 data points for each component, equivalent to a 30-second duration sampled at 100 Hz [Zhu and Beroza, 2019].

These models collectively represent the ongoing drive to leverage the capabilities of deep learning in earthquake detection and phase picking. This offers the potential to enhance the timeliness and accuracy of seismic event identification and analysis. Ultimately, these models each contribute to improved earthquake monitoring and early warning systems.

Table 1: Main earthquake detection ML models. Acronyms: PP - phase picking; PD - phase detection; NC - northern Chile; NCa - northern California; SCa - southern California; STEAD - STanford EArthquake Datasets.

Description	EQT	GPD	PhaseNet	DPP	CRED	BasicPhaseAE
Architecture	CNN-RNN-Attention	CNN	U-Net	CNN/RNN	CNN-RNN	U-Net
Training Data	STEAD	SCa	NCa	NC	SCa	NC
Parameters	376,935	1,741,003	23,305	199,731/546,081	293,569	33,687
Application	PP,PD	PP	PP	PP	PD	PP

3.5 Machine learning toolbox for seismology

To use ML-based detection methods, high quality training data is required. Another requirement is a performance test of each of the above models for the specific data one is applying the models to. SeisBench [Woollam et al., 2022] is a Python-based suite of codes that was developed to compare the performance of the above models, and provide a toolbox for the application of ML algorithms to seismic datasets. The growing number of detection algorithms provided clear motivation for a standardised and unified bench-marking framework for ML detection models.

SeisBench offers an open-source software solution, accessible via GitHub, streamlining access to both machine learning models and datasets. It further simplifies the application of standard data processing and augmentation techniques. This toolbox provides seismologists with an entry point to a diverse array of machine learning models and benchmarking datasets, and encourages community involvement. This ensures its adaptability and relevance in the growing landscape of machine learning within seismology.

4 Machine learning in CO_2 storage monitoring

In the previous section, we explored the applications of ML and DL in earthquake seismology (e.g. see [Mousavi and Beroza, 2023]). This section investigates the potential application of ML and DL in CO_2 storage monitoring.

Yao et al. [2023] conducted an extensive examination of ML applications within the field of CCS. Their investigation revealed that ML algorithms, such as artifical neural networks and convolutional neural networks, were commonly employed for tasks such as predicting physical properties, assessing mechanical stability, and monitoring the migration and leakage of carbon dioxide plumes during carbon dioxide storage. DL methods, including GAN and LSTM, exhibited promising performance in real-time monitoring of carbon dioxide migration and leakage. Decision trees and random forests were primarily employed to establish frameworks for risk assessment and decision analysis, as well as to estimate the success probability of CCS.

Wen et al. [2023] presented the nested Fourier neural operator (Nested FNO), a machine learning approach for dynamic 3D carbon dioxide storage modeling on a basin scale. Figure 14 illustrates the FNO. This method serves as a versatile numerical simulator, suitable for diverse carbon dioxide storage scenarios, including varying reservoir conditions, geological heterogeneity, and injection strategies. The novel deep learning architecture, Fourier neural operator was introduced by Li et al. [2020].

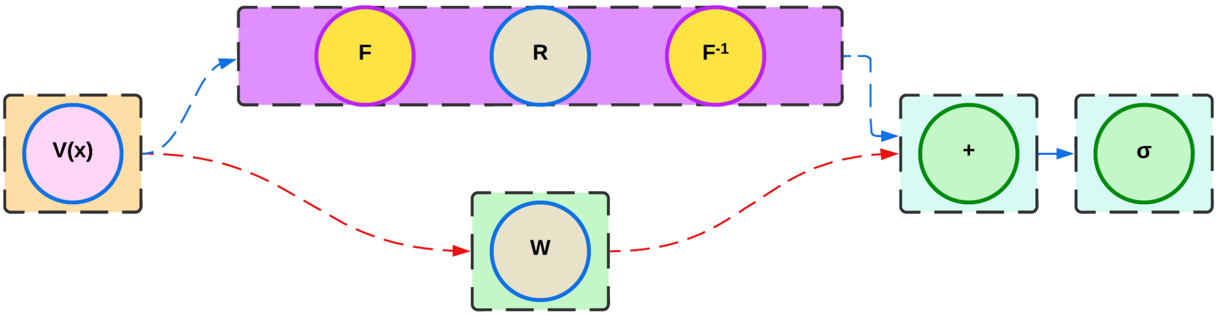

Figure 14: The process of the FNO. The input function V(x) takes two paths through the Fourier layers. In the top path, V(x) undergoes a Fourier transform F, a linear transform R applied to the lower Fourier modes, and an inverse Fourier transform $\mathcal{F}^{-1}$. Meanwhile, in the bottom path, V(x) experiences only a local linear transform W. The outputs of both paths are then combined, and an activation function σ is applied.

4.1 SHARP activities

Within the SHARP project specifically, there are several tasks which could be augmented and potentially enhanced through the application of ML methods. Naturally, the most obvious application is in the detection of earthquakes, used for building earthquake databases. ML techniques may improve the detection capabilities of earthquake monitoring agencies more generally, potentially finding smaller and a greater variety of earthquake signals.

ML-based de-noising algorithms could also improve filtering of waveforms, either an input for ML-based detection or more conventional detection algorithms. This task, particularly for large quantities of data, can be very computationally expensive, requiring repeated transformations of the data. Using ML methods could facilitate faster bulk denoising and make pre-processing of some large continuous datasets now a manageable proposition.

There is likely an opportunity to use the large catalogue of earthquakes and their associated waveforms, produced by work package 2 (Seismology), as a training dataset to study ML detection for North Sea seismicity. Having catalogued the data, it could act as a good foundation for an application of transfer learning, customising a trained model for the application to the North Sea. This would be tuned in terms of the characteristics of both the earthquake and the monitoring networks. With the increasing use of fibre optic DAS systems for passive seismic monitoring, ML-methods will likely be needed in both event detection and denoising.

Many of the tasks in both work package 2 (Seismology) and 5 (Hazards and Risk Quantification) require high quality (e.g., high SNR) waveforms. These tasks include ground motion model calibration, stress drop measurements, and focal mechanism inversion. ML detection routines could also be employed to identify the quality of waveforms, providing another means of automating these analyses.

Similar methods could also be employed in the ground motion characterisation work. Site amplification effects are studied through the measurement of the decay of high frequency energy, as recorded by broadband seismometers. These effects can be difficult to measure, particularly when there is aberrant spikes in noise in the high frequencies. Identifying these noise spikes can be time consuming, requiring time series data to be transformed into the frequency domain, and examining of the spectral ratio of the horizontal and vertical components. This is usually carried out manually. ML methods could be employed to significantly reduce the time taken to identify these unwanted noise spikes, using purely time series data. However, a larger collection of labelled training data would potentially be required for this task.

5 Challenges and future directions

5.1 Challenges in Machine Learning-Based Earthquake Detection

Machine learning, including deep learning, has shown great potential in earthquake detection and other applications, but it faces certain challenges. Firstly, data availability and quality are crucial for training accurate models, and for example access to comprehensive and high-quality hand labelled data can be a limitation [Mousavi and Beroza, 2022].

Second, model interpretability is a significant concern, as understanding why a model makes specific predictions is essential for building trust and making informed decisions. Neural networks continue to be treated mostly as black-box functions, where they map a given input to produce a classification output [Chakraborty et al., 2017, Zhang and Zhu, 2018, Li et al., 2022]. This characterisation arises from the inherent complexity of the models, as they intricately map input data to produce classification outputs without offering readily understandable explanations for their decision-making processes. The challenge, therefore, lies in bridging the gap between the high-level abstraction of neural networks and the need for transparent, interpretable insights to ensure the responsible and accountable deployment of models in real-world applications.

Moreover, scalability and uncertainty quantification are vital, especially in tectonic earthquake monitoring, where timely alerts can be hugely impactful. Uncertainty quantification can be developed in many ways, either numerically or probabilistically [Abdar et al., 2021]. Whilst many earthquake detection algorithms give some kind of qualify score to their signal detections, the exact metrics used and underlying physical phenomena that are being identified are still under-developed. In phase picking ML algorithms, the widths of the activation functions, giving the likelihood of a detection of a specific phase over time, can give some approximation to the precision in the measurement. However, due to the convolutional nature of many of these algorithms, the factors which affect the width of these functions, and thus the pick uncertainty, is complex and requires systematic analysis on specific datasets.

5.2 Future Directions and Emerging Technologies

At present, the majority of deep learning methods applied in seismology primarily focus on constructing models using established input-output pairs – training data of detections are feed in and new detections are made in time series that the model has not seen before. Models are generally being used to forecast outcomes for new, unseen inputs. The nature of the detections being made, or not made, often lack interpretability or explainability [Mousavi and Beroza, 2022]. This is a core challenge in ML more generally, but is critical in seismology, where comlpeteness of catalogues is fundamental to many studies. Explainable AI [Hoffman et al., 2018, Xu et al., 2019, Holzinger et al., 2020] holds promise in making deep learning models more transparent and interpretable, and may aid seismologists in understanding the decision-making processes of these models.

Quantum machine learning is another emerging technology that may outperform classical computers on machine learning tasks [Biamonte et al., 2017, Schuld et al., 2015], improving performance and significantly decreasing training time for large models.

DL seismology is rapidly advancing. The focus remains on incorporating domain knowledge and physical laws, enhancing interpretability, addressing noisy data, and generalising beyond known parameters. Strategies include data augmentation, physics-inspired architectures, and introducing physical constrains [bin Waheed et al., 2021]. This fusion offers exciting prospects for solving seismological challenges, driving innovation and broader applications in the field [Mousavi and Beroza, 2022, 2023].

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

Alankrita Aggarwal, Mamta Mittal, and Gopi Battineni. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1):100004, 2021.

Charu C Aggarwal et al. Neural networks and deep learning. *Springer*, 10(978):3, 2018.

Shamima Ahmed, Muneer M Alshater, Anis El Ammari, and Helmi Hammami. Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*, 61:101646, 2022.

Michael J Ahn and Yu-Che Chen. Artificial intelligence in government: potentials, challenges, and the future. In *The 21st annual international conference on digital government research*, pages 243–252, 2020.

Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

Rex Allen. Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B):S225–S242, 1982.

Rex V Allen. Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, 68(5):1521–1532, 1978.

Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J Aljaaf. A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, pages 3–21, 2020.

Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *electronics*, 8(3): 292, 2019.

Isaac Álvarez, Luz García, Sonia Mota, Guillermo Cortés, Carmen Benítez, and Ángel De la Torre. An automatic p-phase picking algorithm based on adaptive multiband processing. *IEEE Geoscience and remote sensing letters*, 10(6):1488–1492, 2013.

Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74, 2021.

Jafar Alzubi, Anand Nayyar, and Akshi Kumar. Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series*, volume 1142, page 012012. IOP Publishing, 2018.

Denis Anikiev, Claire Birnie, Umair bin Waheed, Tariq Alkhalifah, Chen Gu, Dirk J Verschuur, and Leo Eisner. Machine learning in microseismic monitoring. *Earth-Science Reviews*, page 104371, 2023.

Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, 3:19–48, 2010.

Manfred Baer and Urs Kradolfer. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4):1437–1445, 1987.

Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pages 353–374, 2023.

Andrew G Barto. Reinforcement learning. In *Neural systems for control*, pages 7–30. Elsevier, 1997.

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019.

David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B Tenenbaum, William T Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. *arXiv preprint arXiv:1811.10597*, 2018.

Attila Benko and Cecília Sik Lányi. History of artificial intelligence. In *Encyclopedia of Information Science and Technology, Second Edition*, pages 1759–1762. IGI global, 2009.

Karianne J Bergen, Ting Chen, and Zefeng Li. Preface to the focus section on machine learning in seismology. *Seismological Research Letters*, 90(2A):477–480, 2019.

Gregory C Beroza, Margarita Segou, and S Mostafa Mousavi. Machine learning and earthquake forecasting—next steps. *Nature communications*, 12(1):4761, 2021.

Hetal Bhavsar and Amit Ganatra. A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4):2231–2307, 2012.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549 (7671):195–202, 2017.

Michael J Bianco, Peter Gerstoft, Kim B Olsen, and Fan-Chi Lin. High-resolution seismic tomography of long beach, ca using machine learning. *Scientific reports*, 9(1):14987, 2019.

Umair bin Waheed, Ehsan Haghighat, Tariq Alkhalifah, Chao Song, and Qi Hao. Pinneik: Eikonal solution using physics-informed neural networks. *Computers & Geosciences*, 155:104833, 2021.

Stefano A Bini. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *The Journal of arthroplasty*, 33(8):2358–2361, 2018.

Bogdan Bochenek and Zbigniew Ustrnul. Machine learning in weather prediction and climate analyses—applications and perspectives. *Atmosphere*, 13(2):180, 2022.

Thomas Bornstein, Dietrich Lange, Jannes Münchmeyer, Jack Woollam, Andreas Rietbrock, Grace Barcheck, Ingo Grevemeyer, and Frederik Tilmann. Pickblue: Seismic phase picking for ocean bottom seismometers with deep learning. *arXiv preprint arXiv:2304.06635*, 2023.

Giovanni Briganti and Olivier Le Moine. Artificial intelligence in medicine: today and tomorrow. *Frontiers in medicine*, 7:27, 2020.

Justin R Brown, Gregory C Beroza, and David R Shelly. An autocorrelation method to detect low frequency earthquakes within tremor. *Geophysical Research Letters*, 35(16), 2008.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *arXiv preprint arXiv:1509.00519*, 2015.

Lei Cai, Jingyang Gao, and Di Zhao. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*, 8(11), 2020.

Eduardo Valero Cano, Jubran Akram, and Daniel B Peter. Automatic seismic phase picking based on unsupervised machine-learning classification and content information analysis. *Geophysics*, 86(4):V299–V315, 2021.

Chengping Chai, Monica Maceira, Hector J Santos-Villalobos, Singanallur V Venkatakrishnan, Martin Schoenball, Weiqiang Zhu, Gregory C Beroza, Clifford Thurber, and EGS Collab Team. Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, 47(16):e2020GL088651, 2020.

Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M Rao, et al. Interpretability of deep learning models: A survey of results. In *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*, pages 1–6. IEEE, 2017.

Kit Chambers, J-Michael Kendall, Sverre Brandsberg-Dahl, and Jose Rueda. Testing the ability of surface arrays to monitor microseismic activity. *Geophysical Prospecting*, 58(5):821–830, 2010.

Nitin Kumar Chauhan and Krishna Singh. A review on conventional machine learning vs deep learning. In *2018 International conference on computing, power and communication technologies (GUCON)*, pages 347–352. IEEE, 2018.

Guoyi Chen and Junlun Li. Cubenet: Array-based seismic phase picking with deep learning. *Seismological Society of America*, 93(5):2554–2569, 2022.

Yanbei Chen, Massimiliano Mancini, Xiatian Zhu, and Zeynep Akata. Semi-supervised and unsupervised deep visual learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2022.

Rene Y Choi, Aaron S Coyner, Jayashree Kalpathy-Cramer, Michael F Chiang, and J Peter Campbell. Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology*, 9(2):14–14, 2020.

François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

Sushree Sasmita Dash, Subrat Kumar Nayak, and Debahuti Mishra. A review on machine learning algorithms. *Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 2*, pages 495–507, 2020.

Walter H Delashmit, Michael T Manry, et al. Recent developments in multilayer perceptron neural networks. In *Proceedings of the seventh annual memphis area engineering and science conference, MAESC*, pages 1–15, 2005.

Li Deng. Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA transactions on signal and information processing*, 57:58, 2012.

S Devi, P Malarvezhi, R Dayana, and K Vadivukkarasi. A comprehensive survey on autonomous driving cars: A perspective view. *Wireless Personal Communications*, 114(3):2121–2133, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Rahul Dey and Fathi M Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pages 1597–1600. IEEE, 2017.

Zihan Ding, Yanhua Huang, Hang Yuan, and Hao Dong. Introduction to reinforcement learning. *Deep reinforcement learning: fundamentals, research and applications*, pages 47–123, 2020.

Julian Drew, Robert S White, Frederik Tilmann, and Jon Tarasewicz. Coalescence microseismic mapping. *Geophysical Journal International*, 195(3):1773–1785, 2013.

Xuedan Du, Yinghao Cai, Shuo Wang, and Leijie Zhang. Overview of deep learning. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 159–164. IEEE, 2016.

M Durgadevi et al. Generative adversarial network (gan): a general review on different variants of gan and applications. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pages 1–8. IEEE, 2021.

Mohammed Amine El Mrabet, Khalid El Makkaoui, and Ahmed Faize. Supervised machine learning: A survey. In *2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)*, pages 1–10. IEEE, 2021.

Issam El Naqa and Martin J Murphy. *What is machine learning?* Springer, 2015.

Ngozi Clara Eli-Chukwu. Applications of artificial intelligence in agriculture: A review. *Engineering, Technology & Applied Science Research*, 9 (4), 2019.

Mustafa Ergen. What is artificial intelligence? technical considerations and future perception. *Anatolian J. Cardiol*, 22(2):5–7, 2019.

Mariusz Flasiński and Mariusz Flasiński. History of artificial intelligence. *Introduction to artificial intelligence*, pages 3–13, 2016.

Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G Bellemare, Joelle Pineau, et al. An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, 11(3-4):219–354, 2018.

Rui Fu, Zuo Zhang, and Li Li. Using lstm and gru neural network methods for traffic flow prediction. In *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*, pages 324–328. IEEE, 2016.

Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636, 1998.

Adrian Iustin Georgevici and Marius Terblanche. Neural networks and deep learning: a brief introduction. *Intensive Care Medicine*, 45(5): 712–714, 2019.

Steven J Gibbons and Frode Ringdal. The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1):149–166, 2006.

Valeri G Gitis and Alexander B Derendyaev. Machine learning methods for seismic hazards forecast. *Geosciences*, 9(7):308, 2019.

Pierre Yves Glorennec. Reinforcement learning: An overview. In *Proceedings European Symposium on Intelligent Techniques (ESIT-00), Aachen, Germany*, pages 14–15. Citeseer, 2000.

Liang Gonog and Yimin Zhou. A review: generative adversarial networks. In *2019 14th IEEE conference on industrial electronics and applications (ICIEA)*, pages 505–510. IEEE, 2019.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

Casey S Greene, Jie Tan, Matthew Ung, Jason H Moore, and Chao Cheng. Big data bioinformatics. *Journal of cellular physiology*, 229(12): 1896–1900, 2014.

Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

Chao Guo, Tieyuan Zhu, Yongtao Gao, Shunchuan Wu, and Jian Sun. Aenet: Automatic picking of p-wave first arrivals using deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5293–5303, 2020.

Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

Lav Kumar Gupta, Deepika Koundal, and Shweta Mongia. Explainable methods for image-based deep learning: a review. *Archives of Computational Methods in Engineering*, 30(4):2651–2666, 2023.

Michael Haenlein and Andreas Kaplan. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4):5–14, 2019.

Abid Haleem, Mohd Javaid, Mohd Asim Qadri, Ravi Pratap Singh, and Rajiv Suman. Artificial intelligence (ai) applications for marketing: A literature-based study. *International Journal of Intelligent Networks*, 2022.

Muhammad Salman Haleem. Advances in artificial intelligence, machine learning and deep learning applications. *Electronics*, 12(18):3780, 2023.

Xing Hao, Guigang Zhang, and Shang Ma. Deep learning. *International Journal of Semantic Computing*, 10(03):417–439, 2016.

Ehsan Harirchian, Vandana Kumari, Kirti Jadhav, Rohan Raj Das, Shahla Rasulzade, and Tom Lahmer. A machine learning framework for assessing seismic hazard safety of reinforced concrete buildings. *Applied Sciences*, 10(20):7153, 2020.

Emilia Hattula, Lingli Zhu, Jere Raninen, Juha Oksanen, and Juha Hyyppä. Advantages of using transfer learning technology with a quantative measurement. *Remote Sensing*, 15(17), 2023. ISSN 2072-4292. doi: 10.3390/rs15174278. URL https://www.mdpi.com/2072-4292/15/17/4278.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

J Matthew Helm, Andrew M Swiergosz, Heather S Haeberle, Jaret M Karnuta, Jonathan L Schaffer, Viktor E Krebs, Andrew I Spitzer, and Prem N Ramkumar. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13:69–76, 2020.

Pablo D Hernández, Jaime A Ramírez, and Marcelo A Soto. Deep-learning-based earthquake detection for fiber-optic distributed acoustic sensing. *Journal of Lightwave Technology*, 40(8):2639–2650, 2022.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608*, 2018.

Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. Explainable ai methods-a brief overview. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pages 13–38. Springer, 2020.

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

Mohammadreza Iman, Hamid Reza Arabnia, and Khaled Rasheed. A review of deep transfer learning and recent advancements. *Technologies*, 11(2):40, 2023.

Abdul Jabbar, Xi Li, and Bourahla Omar. A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)*, 54(8):1–49, 2021.

Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, 2021.

Pengcheng Jiao and Amir H Alavi. Artificial intelligence in seismology: advent, performance and future trends. *Geoscience Frontiers*, 11(3): 739–744, 2020.

Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

Manfred Joswig. Pattern recognition for earthquake detection. *Bulletin of the Seismological Society of America*, 80(1):170–186, 1990.

Manfred Joswig. Automated classification of local earthquake data in the bug small array. *Geophysical Journal International*, 120(2):262–286, 1995.

Manfred Joswig and Hartwig Schulte-Theis. Master-event correlations of weak local earthquakes by dynamic waveform matching. *Geophysical Journal International*, 113(3):562–574, 1993.

Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

Juha Karhunen, Tapani Raiko, and KyungHyun Cho. Unsupervised deep learning: A short review. *Advances in independent component analysis and learning machines*, pages 125–142, 2015.

Harpreet Kaur, Qie Zhang, Philipp Witte, Luming Liang, Long Wu, and Sergey Fomel. Deep-learning-based 3d fault detection for carbon capture and storage. *Geophysics*, 88(4):IM101–IM112, 2023.

Aditya Khamparia and Karan Mehtab Singh. A systematic review on deep learning architectures and applications. *Expert Systems*, 36(3): e12400, 2019.

B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2021.

Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

Qingkai Kong, Daniel T Trugman, Zachary E Ross, Michael J Bianco, Brendan J Meade, and Peter Gerstoft. Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1):3–14, 2019.

Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

Ludger Küperkoch, Thomas Meier, and Tobias Diehl. Automated event and phase identification. In *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pages 1–52. Deutsches GeoForschungsZentrum GFZ, 2012.

Nadège Langet, Alessia Maggi, Alberto Michelini, and Florent Brenguier. Continuous kurtosis-based migration for seismic event detection and location, with application to piton de la fournaise volcano, la reunion. *Bulletin of the Seismological Society of America*, 104(1):229–246, 2014.

Sacha Lapins, Berhe Goitom, J-Michael Kendall, Maximilian J Werner, Katharine V Cashman, and James OS Hammond. A little data goes a long way: Automating seismic phase arrival picking at nabro volcano with transfer learning. *Journal of Geophysical Research: Solid Earth*, 126(7):e2021JB021910, 2021.

Pedro Larrañaga, David Atienza, Javier Diaz-Rozo, Alberto Ogbechie, Carlos Esteban Puerto-Santana, and Concha Bielza. *Industrial applications of machine learning*. CRC press, 2018.

M Leonard. Comparison of manual and automatic onset time picking. *Bulletin of the Seismological Society of America*, 90(6):1384–1390, 2000.

Bo-hu Li, Bao-cun Hou, Wen-tao Yu, Xiao-bing Lu, and Chun-wei Yang. Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, 18:86–96, 2017.

Shengbo Eben Li. Deep reinforcement learning. In *Reinforcement Learning for Sequential Decision and Optimal Control*, pages 365–402. Springer, 2023.

Shucai Li, Bin Liu, Yuxiao Ren, Yangkang Chen, Senlin Yang, Yunhai Wang, and Peng Jiang. Deep-learning inversion of seismic data. *arXiv preprint arXiv:1901.07733*, 2019.

Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12):3197–3234, 2022.

Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017.

Zefeng Li, Men-Andrin Meier, Egill Hauksson, Zhongwen Zhan, and Jennifer Andrews. Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, 45(10):4773–4779, 2018.

Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 2021.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.

Bin Liu, Senlin Yang, Yuxiao Ren, Xinji Xu, Peng Jiang, and Yangkang Chen. Deep-learning seismic full-waveform inversion for realistic structural models. *Geophysics*, 86(1):R31–R44, 2021.

Somayya Madakam, Takahiro Uchiya, Shlomo Mark, and Yotam Lurie. Artificial intelligence, machine learning and deep learning (literature: review and metrics). *Asia-Pacific Journal of Management Research and Innovation*, 18(1-2):7–23, 2022.

Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

Mufti Mahmud, Mohammed Shamim Kaiser, Amir Hussain, and Stefano Vassanelli. Applications of deep learning and reinforcement learning to biological data. *IEEE transactions on neural networks and learning systems*, 29(6):2063–2079, 2018.

Dave Martinez, Nick Malyska, Bill Streilein, D Reynolds, William Campbell, F Richardson, C Dagli, C Sahin, V Gadepally, A Tran, et al. Artificial intelligence: Short history, present developments, and future outlook. final report, 2019.

Amitha Mathew, P Amudha, and S Sivakumari. Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, pages 599–608, 2021.

Marian Mazzone and Ahmed Elgammal. Art, creativity, and the potential of artificial intelligence. In *Arts*, volume 8, page 26. MDPI, 2019.

Ian W McBrearty and Gregory C Beroza. Earthquake phase association with graph neural networks. *Bulletin of the Seismological Society of America*, 113(2):524–547, 2023.

Larry R Medsker and LC Jain. Recurrent neural networks. *Design and Applications*, 5(64-67):2, 2001.

Nait Amar Menad, Abdolhossein Hemmati-Sarapardeh, Amir Varamesh, and Shahaboddin Shamshirband. Predicting solubility of co2 in brine by advanced machine learning systems: Application to carbon capture and sequestration. *Journal of CO2 Utilization*, 33:83–95, 2019.

Amr E Mohamed. Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied*, 7 (2):1–15, 2017.

Felix Mohr, Marcel Wever, and Eyke Hüllermeier. Ml-plan: Automated machine learning via hierarchical planning. *Machine Learning*, 107: 1495–1515, 2018.

Claire Monteleoni, Gavin A Schmidt, and Scott McQuade. Climate informatics: accelerating discovering in climate science with machine learning. *Computing in Science & Engineering*, 15(5):32–40, 2013.

Ben Moseley, Tarje Nissen-Meyer, and Andrew Markham. Deep learning for fast simulation of seismic waves in complex media. *Solid Earth*, 11(4):1527–1549, 2020.

S Mostafa Mousavi and Gregory C Beroza. Deep-learning seismology. *Science*, 377(6607):eabm4470, 2022.

S Mostafa Mousavi and Gregory C Beroza. Machine learning in earthquake seismology. *Annual Review of Earth and Planetary Sciences*, 51: 105–129, 2023.

S Mostafa Mousavi, Weiqiang Zhu, Yixiao Sheng, and Gregory C Beroza. Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific reports*, 9(1):10267, 2019.

S Mostafa Mousavi, William L Ellsworth, Weiqiang Zhu, Lindsay Y Chuang, and Gregory C Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1):3952, 2020.

Dawei Mu, En-Jui Lee, and Po Chen. Rapid earthquake detection through gpu-based template matching. *Computers & Geosciences*, 109: 305–314, 2017.

Iqbal Muhammad and Zhu Yan. Supervised machine learning approaches: A survey. *ICTACT Journal on Soft Computing*, 5(3), 2015.

James N Murdock and Charles R Hutt. A new event detector designed for the seismic research observatories. Technical report, US Geological Survey, 1983.

Samreen Naeem, Aqib Ali, Sania Anam, and Muhammad Munawar Ahmed. An unsupervised machine learning algorithms: Comprehensive review. *Int. J. Comput. Digit. Syst*, 2023.

Vladimir Nasteski. An overview of the supervised machine learning methods. *Horizons. b*, 4:51–62, 2017.

Shuteng Niu, Yongxin Liu, Jian Wang, and Houbing Song. A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, 1(2):151–166, 2020.

Salima Omar, Asri Ngadi, and Hamid H Jebur. Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications*, 79(2), 2013.

Pariwat Ongsulee. Artificial intelligence, machine learning and deep learning. In *2017 15th international conference on ICT and knowledge engineering (ICT&KE)*, pages 1–6. IEEE, 2017.

Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

FY Osisanwo, JET Akinsola, O Awodele, JO Hinmikaiye, O Olakanmi, J Akinjobi, et al. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138, 2017.

Daniel W Otter, Julian R Medina, and Jugal K Kalita. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, 32(2):604–624, 2020.

Julio-Omar Palacio-Niño and Fernando Berzal. Evaluation metrics for unsupervised learning algorithms. *arXiv preprint arXiv:1905.05667*, 2019.

Harikumar Pallathadka, Edwin Hernan Ramirez-Asis, Telmo Pablo Loli-Poma, Karthikeyan Kaliyaperumal, Randy Joy Magno Ventayen, and Mohd Naved. Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*, 80: 2610–2613, 2023.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

Darpan Pandey, Kamal Niwaria, and Bharti Chourasia. Machine learning algorithms: A review. *Machine Learning*, 6(02), 2019.

Esteban Pardo, Carmen Garfias, and Norberto Malpica. Seismic phase picking using convolutional networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):7086–7092, 2019.

Walter Hugo Lopez Pinaya, Sandra Vieira, Rafael Garcia-Dias, and Andrea Mechelli. Autoencoders. In *Machine learning*, pages 193–208. Elsevier, 2020.

David I Poole, Randy G Goebel, and Alan K Mackworth. *Computational intelligence*, volume 1. Oxford University Press Oxford, 1998.

Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588, 2009.

Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and Sundaraja S Iyengar. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51(5):1–36, 2018.

M Praveena and V Jaiganesh. A literature review on supervised machine learning algorithms and boosting process. *International Journal of Computer Applications*, 169(8):32–35, 2017.

Wang Qiang and Zhan Zhongli. Reinforcement learning model, algorithms and its application. In *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, pages 1143–1146. IEEE, 2011.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

Puru Rattan, Daniel D Penrice, and Douglas A Simonetto. Artificial intelligence and machine learning: what you always wanted to know but were afraid to ask. *Gastro Hep Advances*, 1(1):70–78, 2022.

Susmita Ray. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE, 2019.

Khalid Raza and Nripendra K Singh. A tour of unsupervised deep learning for medical image analysis. *Current Medical Imaging*, 17(9): 1059–1077, 2021.

Andrew Reynen and Pascal Audet. Supervised machine learning on a network scale: Application to seismic event classification and detection. *Geophysical Journal International*, 210(3):1394–1409, 2017.

David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55(2):1–96, 2022.

Zachary E Ross, Men-Andrin Meier, and Egill Hauksson. P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120–5129, 2018a.

Zachary E Ross, Men-Andrin Meier, Egill Hauksson, and Thomas H Heaton. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901, 2018b.

Zachary E Ross, Yisong Yue, Men-Andrin Meier, Egill Hauksson, and Thomas H Heaton. Phaselink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, 124(1):856–869, 2019.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088): 533–536, 1986.

Nicole Rusk. Deep learning. *Nature Methods*, 13(1):35–35, 2016.

Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.

Iqbal H Sarker. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):420, 2021a.

Iqbal H Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3):160, 2021b.

Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.

Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

Mohit Sewak, Sanjay K Sahay, and Hemant Rathore. An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Nanoscience*, 17(1):182–188, 2020.

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv preprint arXiv:1610.03295*, 2016.

Lihi Shiloh, Avishay Eyal, and Raja Giryes. Efficient processing of distributed acoustic sensing data using a deep learning approach. *Journal of Lightwave Technology*, 37(18):4755–4762, 2019.

Pramila P Shinde and Seema Shah. A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–6. IEEE, 2018.

Hojat Shirmard, Ehsan Farahbakhsh, R Dietmar Müller, and Rohitash Chandra. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268:112750, 2022.

Ajay Shrestha and Ausif Mahmood. Review of deep learning algorithms and architectures. *IEEE access*, 7:53040–53065, 2019.

Amit K Shukla, Manvendra Janmaijaya, Ajith Abraham, and Pranab K Muhuri. Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Engineering Applications of Artificial Intelligence*, 85:517–532, 2019.

Reinaldo Mozart Silva, Lais Baroni, Rodrigo S Ferreira, Daniel Civitarese, Daniela Szwarcman, and Emilio Vital Brazil. Netherlands dataset: A new public dataset for machine learning in seismic interpretation. *arXiv preprint arXiv:1904.00770*, 2019.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. In *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pages 1310–1315. Ieee, 2016.

Jaswinder Singh and Rajdeep Banerjee. A study on single and multi-layer perceptron neural network. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pages 35–40. IEEE, 2019.

Pramod Singh and Pramod Singh. Supervised machine learning. *Learn PySpark: Build Python-Based Machine Learning and Deep Learning Models*, pages 117–159, 2019.

Anirbid Sircar, Kriti Yadav, Kamakshi Rayavarapu, Namrata Bist, and Hemangi Oza. Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 6(4):379–391, 2021.

Hugo Soto and Bernd Schurr. Deepphasepick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks. *Geophysical Journal International*, 227(2):1268–1294, 2021.

Anna L Stork, Alan F Baird, Steve A Horne, Garth Naldrett, Sacha Lapins, J-Michael Kendall, James Wookey, James P Verdon, Andy Clarke, and Anna Williams. Application of machine learning to microseismic event detection in distributed acoustic sensing data. *Geophysics*, 85(5):KS149–KS160, 2020.

Heung-Il Suk. An introduction to neural networks and deep learning. In *Deep Learning for Medical Image Analysis*, pages 3–24. Elsevier, 2017.

Ilya Sutskever. *Training recurrent neural networks*. University of Toronto Toronto, ON, Canada, 2013.

Ilya Sutskever, James Martens, and Geoffrey E Hinton. Generating text with recurrent neural networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 1017–1024, 2011.

Richard S Sutton. A special issue of machine learning on reinforcement learning. *Machine learning*, 8, 1992.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Sudarshana Tamuly, C Jyotsna, and J Amudha. Deep learning model for image classification. In *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*, pages 312–320. Springer, 2020.

Zeeshan Tariq, Murtada Saleh Aljawad, Amjed Hasan, Mobeen Murtaza, Emad Mohammed, Ammar El-Husseiny, Sulaiman A Alarifi, Mohamed Mahmoud, and Abdulazeez Abdulraheem. A systematic review of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum Exploration and Production Technology*, pages 1–36, 2021.

Mohammad Mustafa Taye. Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers*, 12(5):91, 2023.

Matthew E Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(7), 2009.

Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, and Nuwan Kuruwitaarachchi. Real-time credit card fraud detection using machine learning. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 488–493. IEEE, 2019.

Sebastian Thrun and Anton Schwartz. Finding structure in reinforcement learning. *Advances in neural information processing systems*, 7, 1994.

Luca Trani, Giuliano Andrea Pagani, João Paulo Pereira Zanetti, Camille Chapeland, and Läslo Evers. Deepquake—an application of cnn for seismo-acoustic event classification in the netherlands. *Computers & Geosciences*, 159:104980, 2022.

Andreas K Triantafyllidis and Athanasios Tsanas. Applications of machine learning in real-life digital health interventions: review of the literature. *Journal of medical Internet research*, 21(4):e12286, 2019.

Amadej Trnkoczy. Understanding and parameter setting of sta/lta trigger algorithm. In *New manual of seismological observatory practice (NMSOP)*, pages 1–20. Deutsches GeoForschungsZentrum GFZ, 2009.

Alan M Turing. Computing machinery and intelligence. *Mind*, page 433–460, 1950.

MPA Van den Ende and JP Ampuero. Evaluating seismic beamforming capabilities of distributed acoustic sensing arrays, solid earth, 12, 915–934, 2021.

Mirko Van der Baan and Christian Jutten. Neural networks in geophysical applications. *Geophysics*, 65(4):1032–1047, 2000.

Rocio Vargas, Amir Mosavi, and Ramon Ruiz. Deep learning: a review. 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, Eftychios Protopapadakis, et al. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.

Hua Wang, Cuiqin Ma, and Lijuan Zhou. A brief review of machine learning and its application. In *2009 international conference on information engineering and computer science*, pages 1–4. IEEE, 2009.

Jian Wang, Zhuowei Xiao, Chang Liu, Dapeng Zhao, and Zhenxing Yao. Deep learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth*, 124(7):6612–6624, 2019.

Lihui Wang. From intelligence science to intelligent manufacturing. *Engineering*, 5(4):615–618, 2019.

Zhen Wang, Haibin Di, Muhammad Amir Shafiq, Yazeed Alaudah, and Ghassan AlRegib. Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. *The Leading Edge*, 37(6):451–461, 2018.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

Gege Wen, Catherine Hay, and Sally M Benson. Ccsnet: a deep learning modeling suite for co2 storage. *Advances in Water Resources*, 155: 104009, 2021.

Gege Wen, Zongyi Li, Qirui Long, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M Benson. Real-time high-resolution co 2 geological storage prediction using nested fourier neural operators. *Energy & Environmental Science*, 16(4):1732–1741, 2023.

Tom Winder, Conor Bacon, Jonathan Smith, Thomas Hudson, Tim Greenfield, and Robert White. Quakemigrate: a modular, open-source python package for automatic earthquake detection and location. *Authorea Preprints*, 2022.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45, 2020.

Jack Woollam, Andreas Rietbrock, Angel Bueno, and Silvio De Angelis. Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, 90(2A):491–502, 2019.

Jack Woollam, Jannes Münchmeyer, Frederik Tilmann, Andreas Rietbrock, Dietrich Lange, Thomas Bornstein, Tobias Diehl, Carlo Giunchi, Florian Haslinger, Dario Jozinović, et al. Seisbench—a toolbox for machine learning in seismology. *Seismological Research Letters*, 93(3): 1695–1709, 2022.

Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. Explainable ai: A brief survey on history, research areas, approaches and challenges. In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*, pages 563–574. Springer, 2019.

Yongjun Xu, Xin Liu, Xin Cao, Changping Huang, Enke Liu, Sen Qian, Xingchen Liu, Yanjun Wu, Fengliang Dong, Cheng-Wei Qiu, et al. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4), 2021.

Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9:611–629, 2018.

Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer learning*. Cambridge University Press, 2020.

Peiyi Yao, Ziwang Yu, Yanjun Zhang, and Tianfu Xu. Application of machine learning in carbon capture and storage: An in-depth insight from the perspective of geoscience. *Fuel*, 333:126296, 2023.

Clara E Yoon, Ossian O'Reilly, Karianne J Bergen, and Gregory C Beroza. Earthquake detection through computationally efficient similarity search. *Science advances*, 1(11):e1501057, 2015.

Aram You, Jin Kuk Kim, Ik Hee Ryu, and Tae Keun Yoo. Application of generative adversarial networks (gan) for ophthalmology image domains: a survey. *Eye and Vision*, 9(1):1–19, 2022.

Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.

SM Zahraee, M Khalaji Assadi, and R Saidur. Application of artificial intelligence methods for hybrid energy system optimization. *Renewable and sustainable energy reviews*, 66:617–630, 2016.

Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1): 1–27, 2019.

Guoyin Zhang, Zhizhang Wang, and Yangkang Chen. Deep learning for seismic lithology prediction. *Geophysical Journal International*, 215 (2):1368–1387, 2018.

Jian Zhang, Jingye Li, Xiaohong Chen, Yuanqiang Li, Guangtan Huang, and Yangkang Chen. Robust deep learning seismic inversion with a priori initial model constraint. *Geophysical Journal International*, 225(3):2001–2019, 2021.

Quan-shi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39, 2018.

Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE symposium series on computational intelligence (SSCI)*, pages 737–744. IEEE, 2020.

York Zheng, Qie Zhang, Anar Yusifov, and Yunzhi Shi. Applications of supervised deep learning for seismic interpretation and inversion. *The Leading Edge*, 38(7):526–533, 2019.

Weiqiang Zhu and Gregory C Beroza. Phasenet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019.

Weiqiang Zhu, Ettore Biondi, Jiaxuan Li, Jiuxun Yin, Zachary E Ross, and Zhongwen Zhan. Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning. *arXiv preprint arXiv:2302.08747*, 2023a.

Weiqiang Zhu, Alvin Brian Hou, Robert Yang, Avoy Datta, S Mostafa Mousavi, William L Ellsworth, and Gregory C Beroza. Quakeflow: a scalable machine-learning-based earthquake monitoring workflow with cloud computing. *Geophysical Journal International*, 232(1):684–693, 2023b.

Zhuangdi Zhu, Kaixiang Lin, Anil K Jain, and Jiayu Zhou. Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023c.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.