Prestieesci Research Review

# A STUDY TO COMPARE MACHINE LEARNING AND DEEP LEARNING DIFFERENCES

Zhang Yajuan, Midhun chakkaravarthy

Lincoln University College, 47301 Petaling Jaya, Selangor D. E., Malaysia.

Corresponding author: Zhang Yajuan, Lincoln University College, 47301 Petaling Jaya, Selangor D. E., Malaysia, Email: 501743141@qq.com

## ABSTRACT

The revolutionary subfields of computer science known as machine learning and deep learning are finding significant application in the business world. The process of teaching computers and other machines how to make predictions based on prior data or actions using examples from their own memory is known as machine learning. Deep learning is a subsection of machine learning that makes use of artificial neural network techniques and algorithms to train and learn from data that is not structured. This allows for learning to take place from data that is not organized. In order to make sense of the mountain of data that is being created each day, there is an urgent need for techniques of data usage and management that are highly automated and technologically advanced. The software for machine learning (ML) and deep learning (DL) is subjected to a thorough investigation that we provide in this work. The study serves as an introduction to the fundamentals of ML and DL. The most widely used approaches and techniques in fields made feasible by technological advancements are investigated next. In conclusion, a business point of view is presented on the two applications of ML and DL that are most often used.

KEYWORDS: Deep Learning, Machine Learning, Artificial intelligence

## INTRODUCTION

Data science is a rapidly expanding area, and machine learning is becoming an increasingly vital part of it. Through statistically-based data mining projects, students get experience with algorithms for classifying data, making predictions, and unearthing hidden patterns and relationships. Decisions made inside applications and businesses based on these information may ultimately affect key growth KPIs. With the proliferation of big data, it is expected that the need for data scientists would rise.

They will be accountable for aiding in the identification of which business issues are of the biggest relevance and which data is necessary to solve those concerns.

Frameworks like TensorFlow and PyTorch are often used to speed up the creation of machine learning algorithms. The revolutionary new business studio that leverages data mining to fuel cutting-edge generative AI models in addition to more conventional machine learning. Artificial intelligence encompasses a wide range of technologies, including Neural Machines, Learning Machines, and Deep Networks. Although "deep learning" and "machine learning" are often used interchangeably, there are important distinctions between the two that should be made clear. Machine learning, deep learning, and neural networks are only few of the many subfields within artificial intelligence. The area of machine learning that deep learning belongs to is neural networks. The field of deep learning is a subset of neural networks.

While algorithms are fundamental to both deep learning and machine learning, the methodologies used by the two are quite different. While "deep" machine learning might theoretically benefit from labelled datasets (also known as "supervised learning"), it is not required to do so. With access to unprocessed, unstructured data (such as text or images), deep learning may be able to discover on its own which attributes are most useful for distinguishing between the various data types. As a result, more comprehensive datasets may be used with far less intervention from humans. Lex Fridman, in an MIT lecture (link is external to IBM), defines deep learning as "scalable machine learning."

Traditional machine learning, also known as "non-deep" machine learning or "classical" machine learning, relies on human guidance to improve. In most situations, learning requires more organised data, but human experts are still required to build the set of characteristics needed to distinguish different data sources.

Artificial neural networks (ANNs) are built similarly to biological neural networks in that both are hierarchical structures made up of interconnected nodes. A weight and a threshold are given to each artificial neuron, or node, in the network based on its interactions with the other nodes. If the node's output is greater than the preset threshold, then the node will become active on the succeeding network layer. That's the one and only stipulation under which it'll happen. If this condition is not met, the node in question will not send any information to the next layer of the network. "Deep learning" refers to the several interconnected "layers" of a modern neural network, which are essential components of the method. The word "deep" has no depth connotation. To qualify as a deep learning algorithm or a deep neural network, an input and output neural network must contain more than three layers. For simplicity's sake, a neural network wouldn't have more than three layers.

Research in areas like computer vision, natural language processing, and voice recognition has been greatly aided by deep learning and neural networks. Here is where these innovations have had the most significant impact. Read "Artificial Intelligence, Machine Learning, Deep Learning, and Neural Networks: What's the Difference?" to find out more about the connections between these ideas. This article explores the connections between these ideas.

## BACKGROUND OF THE STUDY

Arthur Samuel is said to have used the phrase "machine learning" for the first time in 1952. The perceptron was created by Frank Rosenbelt in 1957 at the Cornell Aeronautical Laboratory. He built on the work of Donald Hebb and Arthur Samuel. In its early stages, the perceptron was thought of as a real device rather than a computer code. The programme was put on the Mark 1 perceptron, a computer that was made just for the job. There was a plan for this perceptron to be used for picture recognition on the IBM 704. Because of this, the software and methods could be sent to other computers and used there instead. Many people think that the closest friend algorithm, which was created in 1967, was the first step towards basic pattern recognition. Being one of the first programmes to try to figure out the best route for visiting salesmen, this method was helpful. Road maps were made with it, and it was one of the first systems to do so. It was used by the seller to find good (but not always the best) ways to get to the place they chose. While there was some progress in the 1950s and 1960s, it wasn't until the late 1970s that there was a lot of real progress. Several things led to this, but the most important one was how famous Von Neumann architecture was. A lot of people have made software that use this design, which is said to be easier to understand than a neural network because it stores commands and data in the same memory. This framework has been used by many people to make programmes. Still, John Hopfield came up with the idea of building a network of lines that go both ways in 1982. There are some similarities between this and how neurons do their jobs, and this is a common way to use deep learning in the twenty-first century. Furthermore, Japan declared in 1982 that it would be focused on developing more advanced neural networks. This led to funding from the US, which in turn led to more study being done in this area. In the late 1980s and early 1990s, there wasn't much progress in the field of machine learning. Terrence Sejnowski created NETtalk in 1985, a piece of software that takes text as input and compares phonetic transcriptions to learn how to pronounce written English text. Backpropagation was added to neural networks to make them better in 1986, and Yann LeCun created the convolutional neural network in 1989, which included backpropagation. However, IBM's Deep Blue, a computer that plays chess by itself, was made in 1997. In a regular chess match with time limits, it was the first time a machine beat the current world champion. This was seen as an example of a machine being smarter than a person. However, since the beginning of the 21st century, some businesses have seen the value in machine learning and have started to put a lot of money into it to stay ahead of their rivals. A lot of study and development is being done in the area of machine learning because it is becoming more popular (Keith D. Foote, 2019).

## PROBLEM STATEMENT

"Deep learning requires millions of data points whereas machine learning just needs hundreds. Typically, machine learning algorithms do well with smaller datasets. In order to comprehend and outperform conventional machine learning algorithms, Deep Learning needs massive volumes of data."

To develop goal-oriented behaviour and an artificial personality through training and education, AI researchers study neural-like elements and multidimensional neural-like expanding networks, transient and long-term memory, and the functional organisation of the brain of artificial intelligent systems. Artificial intelligence (AI) is a wide term encompassing many different approaches to the use of computers to solve problems via the use of logic, processes, and algorithms. The notion of artificial intelligence has been utilised in programmes for natural language processing, information processing, automated programming, robotics, scenario analysis, game playing, intelligent systems, and scientific theorem proving.

Machine learning is a subfield of AI that enables self-improvement in computers without further manual programming. Algorithms and models of neural networks are used to train computers to become better over time. The goal of machine learning is to develop automated systems capable of analysing data and adapting to new circumstances without human intervention. In order to improve in the future, given the examples we provide, learning begins with observations of data, including direct experience or teaching. In machine learning, algorithms are 'trained' to learn how to identify patterns and features in big datasets and then use that information to make judgements and predictions.

Deep Learning is a subfield of machine learning that computes an output based on an input dataset (metadata) that has been exposed to numerous layers of nonlinear alteration. This approach was developed for use with Artificial Neural Networks (ANNs) and is referred to as Deep Learning (DL). It is a reference to the process of machine learning, in which computers gain information via the process of trial and error as well as statistical analysis without any involvement from a human being. Because of the one-of-a-kind capability of this technique to do automated function extraction, only the characteristics that are absolutely necessary for finding a solution to the problem are retrieved by hand.

## LITERATURE REVIEW

The ability to speak many languages has fascinated scholars since the 17th century. Since then, global languages and mechanical dictionaries have been created to bridge linguistic gaps. Due to the emergence of increasingly internationalised enterprises and globalisation, the capacity to translate texts from one language to another automatically without human intervention has been debated for almost sixty years. Due of this growth, this issue has gained more attention in the recent decade. If a corporation wants to promote its goods globally, they must provide accurate and multilingual product documentation. Complex products that serve administrators, users, and developers may have several papers. For exporters, it may be difficult to locate experienced translators with the technical knowledge to properly translate technical content and provide it at affordable pricing. As a consequence, apprehensive organisations are seeking machine translation solutions. High-quality automatic translations of texts are essential to guarantee that people from different languages can access the same information. Technical documentation is used to provide a steady supply of work, satisfy customers, and describe how things work, therefore firms value precise translations. Inaccuracies here might have serious consequences. Another hurdle to inter-business communication is miscommunication caused by inadequate translations. Because natural language processing is so complicated, machine translation software must be examined before it is judged high-quality. Since the issue cannot be solved by translating formal business papers with software, the focus must shift from document generation to document evaluation and repair.

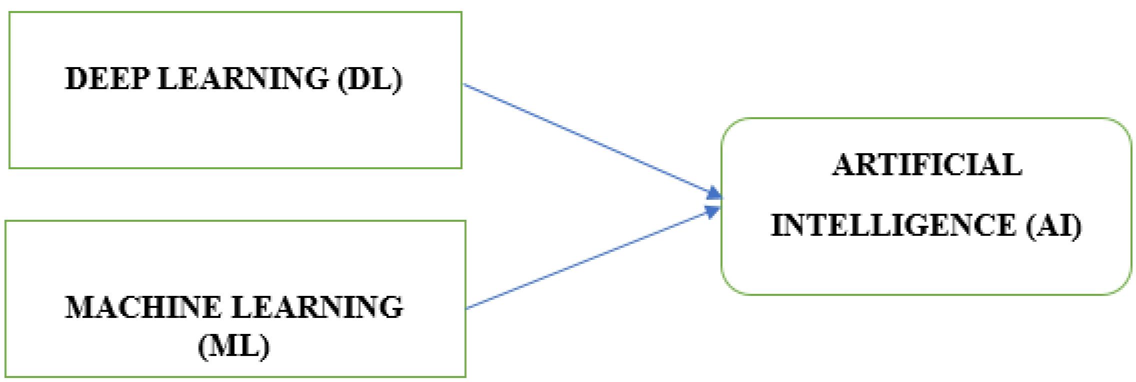
Thus, firms must examine translated technical material to save money, time, and establish a reliable system for translating crucial papers. This maintains a quality foundation. Because "quality" is subjective and has many aspects, it is hard to assess a translation's grammatical correctness, stylistic flourishes, and semantic accuracy. Universal machine translation that can accurately duplicate any statement is unlikely since it is difficult to teach a computer the meaning of a word. This is because teaching a computer a phrase's meaning is hard. To confirm a document's translation into a language that was effectively transmitted to its target audience, it's necessary to assess its quality. This content category is being automatically translated for several reasons, including the massive volume of technical documentation published by every commodity seller. Outsourcing technical document translation to foreign translators helps modern companies address the age-old challenge. The individual who requested the translation may not speak the target language, thus a human translator must do it accurately and professionally. This thesis builds on that by using machine learning to create an algorithm that can tell whether a document was translated by people or robots. This method can distinguish human and machine translations of the same text. This algorithm assembles the basic elements of a document evaluation mechanism.

When undertaking data mining, find the strategy that will solve the problem best. This is commonly done using machine learning methods. Since a long time, humans have had an advantage over computers in that we naturally improve at problem-solving. Humans may evaluate their previous actions and adjust their direction based on new knowledge. Past computers couldn't learn from their mistakes since they didn't consider the implications. Machine learning teaches computers to improve by acquiring, evaluating, and using more data and experience. The first scientist to design self-learning was A. Samuel in 1952. His code improved with each checkers game. The first pattern recognition system detected data patterns in 1967. The computer searched for similarities between freshly obtained and stored data. Since the 1990s, machine learning has been used in data mining, adaptive software systems, and language and text learning. A computer code that learns from its environment and improves its ads for online consumers is dangerously close to becoming smart. Machine learning systems are also categorised by their learning approach. Classification is usually based on algorithm inference.

## RESEARCH OBJECTIVE

1. To find out the difference between deep learning and machine learning.
2. To determine the uses of deep learning and machine learning.
3. To identify the differentiate machine learning from deep learning and deep learning from artificial intelligence.

## CONSEPTUAL FRAMEWORK




## RESEARCH METHODOLOGY

The goal of quantitative research to find statistically significant relationships between variables by collecting numerical data on those variables and feeding it into statistical models. Quantitative studies aim to get a more in-depth understanding of society. Researchers often use quantitative methods when examining phenomena with a personal effect. Quantitative studies provide hard data in the form of tables and graphs. Quantitative study relies heavily on numerical data, which necessitates a methodical strategy to collecting and analysing the data. It may be used in a variety of ways, including averaging out data, making forecasts, looking into connections, and extrapolating results to bigger populations. Quantitative studies are the polar opposite of qualitative studies, which rely on in-depth interviews and observations. Quantitative research techniques are widely used in many academic disciplines, including biology, chemistry, psychology, economics, sociology, marketing, and many more.

**Sampling:** A pilot study was conducted with the questionnaire using a group final study was conducted with the questionnaire. A total of questionnaires was distributed among customers selected in a systematic random sampling. All the completed questionnaires were considered for the study and any incomplete questionnaire will be rejected by the researcher.

**Data and Measurement:** Primary data for the research study was collected through questionnaire survey. The questionnaire was divided into two parts – (A) Demographic information (B) Factor responses in 5-point Likert Scale for both the online and non-online channels. Secondary data was collected from multiple sources, primarily internet resources.

**Statistical Software:** MS-Excel and SPSS 25 will be used for Statistical analysis.

## RESULT

- **Model Information**

The data has problems with missing data, as seen in Tables 3-4 and 3-5. In addition to the missing data that applies to both alumni and non-alumni, the non-alumni data has a larger proportion of missing data than the alumni data, notably in the education eld. The model focuses on alumni data, and the non-alumni data will be explored independently in future research because of the larger proportion of missing data in the non-alumni data. As was said before, the benefits of Deep Neural Network (DNN) become more apparent with a larger training dataset. Therefore, as was indicated, the ecacy of DNN may be aacted by lowering the quantity of training data. There are both static and dynamic characteristics within the UVic alum data set. Census data (average income, average home worth, and average debts of the constituent's dispersion area), together with demographic information (age, sex, occupation, city, title, and education), are examples of data that will not change over time. During the model's xed time period, such details don't change. Donation information is broken down each year over the last decade, beginning in 2010. Records before to 2010 are legacy data kept in an outdated format that lacks organisation and consistency. There are gaps or inaccuracies in our understanding of things like event participation, political affiliation, responses to appeals, and family dynamics. Time-varying information, such as appeals and campaigns, has been collected more regularly since 2010, whereas donation data has been documented since 1987. Because of this, we restricted the model's usage of time-varying data to events after 2010. The original dataset for feature selection contains 84 features, some of which are categorical dummy variables. When choosing features for the models, the greatest donation amount is used in the feature selection procedure. Chapter 3 delves into the approach and features used for feature selection. There are four groups represented in multi-class models.

Class 1 - $0

More than $1 but less than $1,000 falls into Class 2.

More than $1,000 but less than $10,000 falls under Class 3.

Class 4: $10,000 or more

| Type | Total Count | % |
|---|---|---|
| Non-Alumni | 48,359 | 100% |
| Education Missing or Unknown | 32,989 | 68% |
| Education Known | 15,370 | 32% |
| Age Missing | 33,826 | 70% |
| Age Known | 14,533 | 30% |
| Work Information Missing | 48,204 | 99.7% |
| Work Information Known | 155 | 0.3% |
| No LinkedIn Information Available | - | - |
| Marital Status Missing | 38,945 | 81% |
| Marital Status Known | 9,414 | 19% |
| Family Information Missing | 38,381 | 79% |
| Family Information Known | 9,978 | 21% |

**Table 5: Missing Data Among live Non-alumni**

- **Measures for Success**

Common measures for gauging machine learning effectiveness include accuracy, recall, precision, and F1 scores. The ratio of accurately predicted observations to the total number of observations is the most intuitive metric for assessment. If the actual class and predicted class agree for TP and TN but the actual class and projected class disagree for FP and FN, then the accuracy formula is (TP+TN)/(TP+TN+FP+FN).

That was a definite "Yes"

"True negative" = "TN"

False positive

A false negative result.

While accuracy is often used as a measure of a model's efficacy, it may not necessarily be informative. Some of the issues with precision are:

• False positives and false negatives are also counted as mistakes when evaluating accuracy.

• Accuracy is a useful statistic only when the dataset is balanced, with about equal numbers of false positives and false negatives. The ratio of true positive predictions to false positive predictions is called precision, while the ratio of true positive predictions to all real positive observations is called recall. Last but not least, the F1 score is a weighted average of the accuracy and recall scores. To provide a fair comparison between the standard and our bimodal model, we modified these four assessment criteria.

Preciseness equals (TP+TN) / (TP+FP+TN+FN)

TP / (TP+FP) = Accuracy

The formula for recall is: TP / (TP + FN)

To calculate your F1 Score: Recall * Accuracy / Accuracy

In their experiments, the models tried:

• A Time-Independent Model for Multiclass Neural Networks

• SMOTE Multiclass Time-invariant Neural Network

• Unchanging throughout time Classification using Multi-Support Vector Machine

• Stacked Multiclass LSTM with Time Variation

• Multiclass stacked GRU with time-varying input

• Bidirectional Time-varying Multiclass LSTM

• Bidirectional Time-Varying Multiclass GRU

• Multiclass Long Short-Term Memory Time-Distributed

• Multiclass GRU with Time-Dependent Distributions

• Multi-class recurrent neural network (RNN) with time-invariant conditions (bimodal Model)

Variable in time Time-Invariant Multiclass GRU (bimodal Model)

Please take note that SMOTE refers to the Synthetic Minority Oversampling Method. Tables 4.2–4.5 provide confusion metrics for the chosen models across all tests, and table 4.6 details the overall recall rate by summing the recall rates for classes 1, 2, and 3. This is the number of instances when the minority class was correctly predicted. Overall, the expected recall rate is larger than zero, based on the real 1, 2, and 3 values.

The diagnostic performance charts for the chosen sequential models are shown in Figures 4.1–4.4. They are used in counting epochs for sequential models. One of the hyper-parameters that determines the number of iterations across the training dataset during which the model parameters are updated is the number of epochs. As can be seen in Figures 4.1 and 4.2, the validation loss begins to rise at roughly Epoch 100. Early on, the Distributed model's validation time loss decreases dramatically. Training continues getting better and better while validation loss remains constant. The LSTM with conditions model stops learning as soon as the validation loss decreases. The best epoch number for each model is determined by comparing many possible values.

## CONCLUSION

In order to regularise neural networks, this chapter presented a training method referred known as dense-sparse-dense, or DSD for short. DSD involves first pruning neural network connections, and then rebuilding them thereafter. The abbreviation "DSD" stands for "dense-sparse-dense." Learning which connections are the most crucial in the beginning stages of the intensive instruction for our approach is one of the primary focuses of our attention. Following this step, DSD will "regularise" the network by eliminating any unnecessary connections and retraining to a sparser, more stable solution that will either maintain or improve accuracy. Following this, the network is retrained from the bottom up utilising the patched connections that were generated all during the phase of pruning. As a direct consequence of this fact, the range of application of the model as well as the number of possible dimensions that its parameters might assume are both expanded. The accuracy of the predictions is shown to improve as a direct consequence of DSD training. DSD training is able to dramatically improve the accuracy of CNNs, RNNs, and LSTMs, as shown by our studies on ImageNet using GoogleNet, VGGNet, and ResNet; on Flickr-8K using NeuralTalk; and on the WSJ dataset using DeepSpeech and DeepSpeech-2. These findings are a consequence of the tests that we conducted. In addition to this, we conducted a T-test in order to confirm the importance of the significant advances that were made by DSD training. The findings of the trials indicate that participating in DSD training is good for increasing one's accuracy.

Within the context of this master's thesis, we investigated deep learning architectures for datasets of medium and small sizes pertaining to picture categorization. In the first chapter, we explored the remarkable accuracy of convolutional neural networks as well as their operation and how it works. In the next chapter, we will show how successful the Fine-Tuning approach is by applying it to this particular data set. In addition to this, we spoke into depth about how our bootstrapped version of InceptionV3 ended up winning the DSG online competition. In the last chapter, we went over the advantages and disadvantages of a number of different Weakly Supervised Learning approaches, such as Multi Instance Learning (MIL) and Spatial Transformer Networks (STN). Fine Tuning was also used to the improvement of the Weldon model, which is a particular kind of MIL model.

In a word, artificial intelligence (AI) is a system that provides us with the capacity to often develop spectacular insights from boring data mountains. This ability is made possible by AI. This study reveals that not all methods of artificial intelligence (AI), such as deep learning and machine learning, are appropriate for resolving all types of problems. In spite of the fact that deep learning is a more recent and, presumably, more sophisticated method, decision tree-based models performed far better than deep learning did in this research when it came to predicting the deterioration of land. This was the case even though deep learning is a more current methodology. This demonstrates that while deep learning is beneficial for many applications, more traditional machine learning approaches are still the better option for anticipating land degradation due to their greater accuracy. This is because deep learning requires more data and takes longer to process. In spite of the fact that deep learning has gained increasing attention in recent years, this remains the case.

## LIMITATION

The main goal of quantitative study is to find the numbers in data. A big sample size is generally needed for quantitative study methods. However, this large-scale study can't be done because of a lack of funding. In the future, this can be done by using a bigger group number and taking longer to do this work. The organised questions used in quantitative research can only give limited results, and the results don't always show what really happened in a broad sense. Additionally, quantitative research is hard, costly, and takes a substantial amount of time to analyse. In future studies, the reasons given for why the places in question were not covered in this one will be explained. List the likely problems that other researchers might run into when they are planning future study on the subject. This will help you make a research plan that you can actually carry out, including the study's topic, goal, and method.

In this day and age of Big Data, when photos of any kind and capacities that can be computed are more easily available than at any other time in history, the development, needs, and expectations of Deep Learning are apparent. In this context, the Convolutional Neural Network is the statistical model that has shown to be the most successful when used to picture recognition.

They have access to a large amount of data, which allows them to predict your preferences in advance. On the other hand, it has a number of limitations, which ultimately lead to the creation of deep learning.

## REFERENCES

1. Steenhuis, H.-J. and Pretorius, L. (2017), “The additive manufacturing innovation: A range of implications”, Journal of Manufacturing Technology Management, Vol. 28 No. 1, pp. 122–143.
2. Taylor, A. (2017), “Perspectives on the university as a business: The corporate management structure, neoliberalism and higher education”, Journal for Critical Education Policy Studies, Vol. 15 No. 1, p. 117.
3. Wilson, C., Tyfield, D., 2018. Critical perspectives on disruptive innovation and energy transformation. Energy Res. Soc. Sci. 37, 211–215.
4. Zhang, Han X and Deng C, (2018) Review on the research and practice of deep learning and reinforcement learning in smart grids, CSEE Journal of Power and Energy Systems, vol. 4, no. 3, pp. 362-370.