**PART I: AI Deep Learning (20 Points)**

**Question 1.1:** Provide an overview (at a minimum of 2 pages, including images) of the history of artificial intelligence, including its sub-fields, machine learning, and deep learning.

**Introduction:**

Artificial Intelligence is nothing by the decision-making and response process based on the collection and analysis of data. AI is the common buzzword synonymous with progress that has both been accepted and questioned for its benefits and implications. It's a popular term that represents how technology is advancing. People have been aware of AI for about 70 years. A significant person in AI history is Alan Turing. In 1951, he wrote an article about machines thinking like humans. He devised a test, called the Turing Test, to see if a machine can think like a human.

This idea started a big change in the way we live and work. Even if the changes seem to be happening slowly, AI has made a big difference in our world.

A diagram of a graph

Description automatically generated

Fig: AI Advancements and Boom Period

The history of artificial intelligence (AI) dates back to the mid-20th century when computer scientists began developing algorithms and techniques to simulate human reasoning and intelligence.

**History:**

A graph with a line and text

Description automatically generated with medium confidenceFig: AI Advancements and History Period

**The Starting Point:**

* Birth of AI happened in Dartmouth College, 1956

The term "artificial intelligence" was coined during a historic conference at Dartmouth College in 1956. This event, led by luminaries like John McCarthy, marked the official beginning of AI as a dedicated field of research.

* Early Rule-Based AI (1950s-1960s)

They focused on rule-based systems, which used logical rules to simulate human reasoning. These systems laid the foundation for further AI exploration.

Machine Learning Emerges (1960s-1970s)

* In the 1960s and 1970s, researchers started developing machine learning algorithms that allowed computers to learn from data. This marked a significant shift in AI, enabling systems to improve their performance over time through experience.
* Expert Systems Revolution (1980s)

The 1980s witnessed the rise of expert systems, AI programs capable of emulating human expertise in specific domains. Notable examples include MYCIN for medical diagnosis and Dendral for chemical analysis.

* The AI Winter (1980s-1990s)

AI community faced a period known as the "AI winter." Progress was slower than anticipated, and funding for AI research dwindled. Expectations had to be tempered with the realities of the field.

* Neural Networks Resurge (1990s-2000s)

In the late 20th century, AI research saw a resurgence with the renewed exploration of neural networks. These networks, inspired by the structure of the human brain, offered promise in tackling complex problems.

* Deep Learning Revolution (2010s-Present)

The breakthrough moment for AI came with deep learning. Deep neural networks, with multiple layers, ushered in remarkable advancements in various domains.

* Computer Vision

Convolutional Neural Networks (CNNs) revolutionized computer vision, enabling machines to excel in tasks such as image recognition, object detection, and facial recognition.

* CNN

**Convolutional Neural Network**

A diagram of a network

Description automatically generated

Fig: AI Convolutional CNN Neural Network Period

* Natural Language Processing

Transformers, like BERT and GPT, harnessed the power of attention mechanisms, propelling AI to new heights in natural language understanding and generation.

A diagram of a software program

Description automatically generated

* Reinforcement Learning
* Reinforcement learning, exemplified by AlphaGo and OpenAI's DOTA 2 bot, showcased AI's ability to master complex tasks through trial and error.

**References:**

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine, 27(4), 12-14.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

<https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

Lori Perri, August 17, 2023, on the 4 Trends That Prevail on the Gartner Hype Cycle for AI, 2021

<https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle>

**Question 1.2:** Provide an overview (at a minimum of 1.5 pages, including images) of deep learning, including (but not limited to) the relationship between deep learning and machine learning, artificial intelligence.

**Overview on Deep Learning:**

Deep Learning, a subset of Machine Learning, is inspired by the structure and function of the human brain. It centers on Artificial Neural Networks (ANNs) with multiple interconnected layers. These layers, known as hidden layers, enable the network to automatically learn hierarchical representations of data. Each layer processes information and passes it on to the next, culminating in a final output.

**Relationship Between Deep Learning, Machine Learning, and AI:**

The origination of AI kick-started the development of Machine Learning and then Deep Learning.

Artificial Intelligence is the ideology about how computers think and act like humans. Machine Learning is the task of enabling computers to perform actions and Deep Learning is a subset of Machine learning that is based on artificial neural networks.

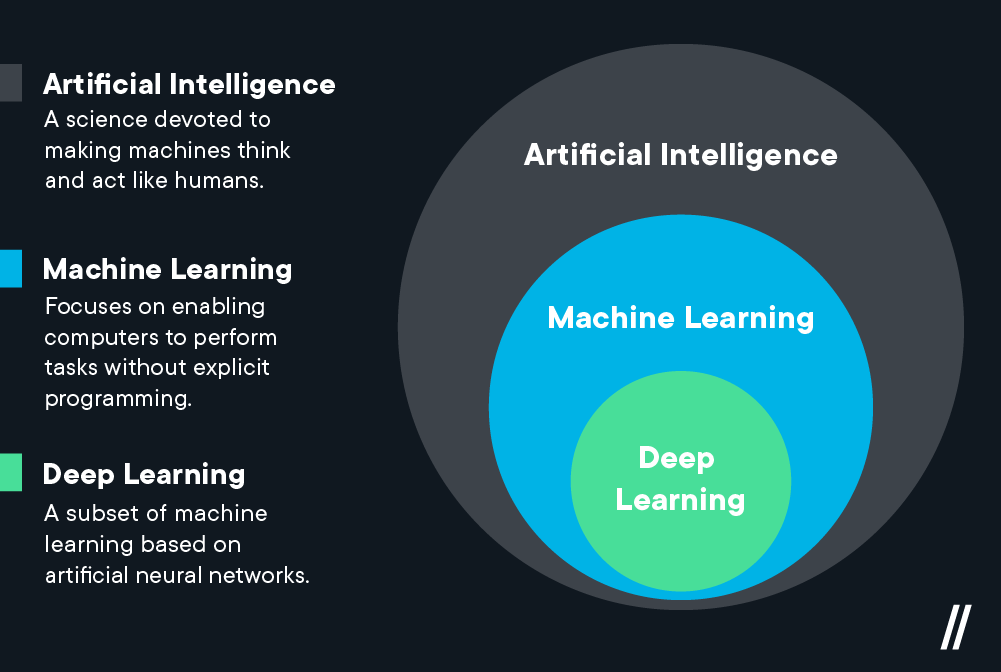


Fig: Relationship of Deep Learning with ML and AI

Machine Learning allows for computers to decipher information step-by-step method. This allows for low computational power while Deep Learning assimilates all inputs in one go and processes information at a higher level by looking at patterns and predictions and determining the best-suited outcome.

|  |  |
| --- | --- |
| Deep Learning | Machine Learning |
| Deep Learning is a subset of Machine Learning, which, in turn, is a subset of Artificial Intelligence. | Machine Learning encompasses various techniques for enabling machines to learn from data, adapt to new information, and make predictions or decisions |
| Deep Learning stands out due to its use of deep neural networks, which are particularly effective at handling complex, high-dimensional data. |  |

**Deep Learning and Artificial Intelligence**

|  |  |
| --- | --- |
| Artificial Intelligence | Deep Learning |
| Artificial Intelligence is the broader field that encompasses both Machine Learning and Deep Learning | Deep Learning has played a pivotal role in AI's progress by enabling machines to process and understand data in ways that resemble human cognitive processes. |
| AI aims to create systems capable of simulating human intelligence, including reasoning, problem-solving, and perception. |  |

Key Concepts in Deep Learning

A diagram of a network

Description automatically generated

Fig: AI Key Concepts in Deep Learning

**Artificial Neural Networks (ANNs)**

Artificial Neural Networks are the foundation of Deep Learning. They consist of layers of interconnected nodes (neurons), each performing mathematical operations on incoming data. ANNs are classified into various types, including feedforward, recurrent, and convolutional networks, tailored for specific tasks.

**Training and Backpropagation**

Deep Learning models are trained using vast datasets. During training, the model learns to adjust its internal parameters to minimize the difference between its predictions and the actual target values. Backpropagation, a crucial technique, is used to update the network's weights based on the error calculated during training.

**Deep Learning Architectures**

Different deep learning architectures are designed for specific tasks:

* Convolutional Neural Networks (CNNs) are primarily used for computer vision tasks, such as image classification and object detection.
* Recurrent Neural Networks (RNNs) are suited for sequential data, making them ideal for natural language processing, speech recognition, and time series analysis.
* Transformers, a newer architecture, excel in tasks involving attention mechanisms and have revolutionized NLP.

**References:**

Michael Middleton, Feb 08, 2021, on Deep Learning vs. Machine Learning, Link: <https://flatironschool.com/blog/deep-learning-vs-machine-learning/>

Rancho Labs, July 14, on 6 Major Sub-Fields of Artificial Intelligence, Link: <https://rancholabs.medium.com/6-major-sub-fields-of-artificial-intelligence-77f6a5b28109>

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep Learning (Vol. 1). MIT press Cambridge.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

**Question 1.3:** Explain (at a minimum of 1.5 pages, including images) why Deep Learning is very popular in recent years.

Deep learning is a specialized method within machine learning, a broader AI domain. It excels at interpreting complex data, from recognizing images to processing speech, but AI encompasses much more, like simulating cell biology to engineering space missions. While deep learning has made significant strides, suggesting it can fully automate diverse tasks like cell replication or space travel without human input is an overstatement. AI, including deep learning, is rapidly advancing but remains a tool that augments human expertise rather than replacing it across all areas.

**The following way of deep learning proves more popular than any other AI technique:**

* Understanding human behavior and attributes: Voice recognition, facial recognition, body movements, sign identification as well as replication into robots have proven robotics as a replacement for man to conduct easy and monotonous as well as difficult and dangerous jobs.
* Identification, classification as well as prediction capability: While Machine learning can enable one of these actions, deep learning can generate the best possible outcome by combining all mentioned techniques. This allows for quick processing and study in fields that require deep analysis and have less room for error.
* Data Availability and Computation Power: With the availability of a tremendous amount of data, we now can structure and use the data available to make informed decisions and in turn collect more data. The computation power for a program is now very high which provides immediate results.
* Some of the applications of Deep Learning are as below:
* Medical Diagnosis – Diagnosis of diseases, prediction of cell patterns and changes.
* Financial management – Identification of credit frauds, stock movements, investment patterns
* Consumer industry – Self-driving cars, human-less stores, consumer behaviors analysis on buying patterns to determine production.
* Security Sector – Judicial decisions based on facial patterns, and ticketing for violations.

Image Recognition: Convolutional Neural Networks (CNNs), a type of deep neural network, have achieved human-level performance in image recognition tasks. Applications include facial recognition, object detection, and medical image analysis.

A diagram of a diagram of a diagram

Description automatically generated

Fig: AI Convolutional CNN Neural Network Period

**Natural Language Processing (NLP)**

Transformers, another class of deep learning models, have revolutionized NLP by enabling machines to understand and generate human language. This has led to advancements in machine translation, chatbots, and sentiment analysis.

A diagram of a diagram of a deep learning application

Description automatically generated with medium confidence

Fig: Example of DL Applications

**References:**

Michael Middleton, Feb 08, 2021, on Deep Learning vs. Machine Learning, Link: <https://flatironschool.com/blog/deep-learning-vs-machine-learning/>

Raja Mitra, April 16, 2018, on Applications of Deep Learning & Big Data Analytics in Organizations, Link: <https://montouche.medium.com/deeplearning-bigdata-applications-22ab73b3163f>

**PART II: MLPs (Fully Connected Neural Networks) with Keras (50 Points)**

1. **MLP Visualization: 8 input variables + 5 neurons + 5 neurons + 2 outputs**

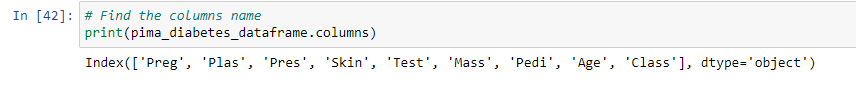
A diagram of a network architecture

Description automatically generated

1. **Database Introduction**

The Pima Diabetes.csv dataset was created by the National Institute of Diabetes and Digestive and Kidney Diseases to diagnostically predict and classify if a patient has diabetes or not patient has diabetes.

The following are the variables included with the dataset that impact the diabetes prediction:



Preg: Number of times pregnant (preg)

Plas: Plasma glucose concentration 2 hours in an oral glucose tolerance test (plas)

Pres: Diastolic blood pressure in mm Hg (pres)

Skin: Triceps skinfold thickness in mm (skin)

Test: 2-Hour serum insulin in mu U/ml (insu)

Mass: Body mass index measured as weight in kg/(height in m)^2 (mass)

Pedi: Diabetes pedigree function (pedi)

Age: Age in years (age)

o class: 0 or 1 (no or yes) (negative or positive). This is the outcome variable that.

predicts if the patient has diabetes or not a patient has diabetes.

1. **Data Preprocessing**

The following preprocessing steps were undertaken to build the initial MLP model in the Jupyter notebook:

It was observed that there were Null values within the observations that would have rendered incorrect calculations while building the model. And hence, the dataset was cleaned as follows:

* Observations that had 0 or Null values were removed.
* The mean and standard deviation of the Insulin variable (test) depicted extreme outliers. Hence the 1% of the outliers were removed to bring the mean lower than the initial.

Post the above-mentioned cleanup, the final number of observations was 768 with 9 variables.

The cleaned dataset was then uploaded into the remote server using the GCP terminal and then loaded in the pima\_diabetes\_dataframe.

1. **Exploratory Data Analysis:**

The dataset consists of 768 observations and 9 variables including the final classification (outcome) variable of patients who have diabetes.

* The blood pressure and skin thickness variables are normally distributed whereas the insulin and age are skewed to the left. The dataset has more observations of patients in age groups less than 25 and not pregnant.

A group of blue and white graphs

Description automatically generated

Fig: Histogram

For example, the 'Preg' histogram shows that more individuals have fewer pregnancies, with the number decreasing as the number of pregnancies increases. Similarly, 'Age' shows that there are more younger individuals in the dataset and fewer older individuals.

overall shape of the data, detecting outliers, and identifying the distribution of the variables, which is critical for data preprocessing with statistical methods or machine learning algorithms for analysis.

1. **Model Building**

Sequential Modeling was applied to build the model. The dataset was split into 70% train and 30% test allowing for random seeding for a generation.

The model built consisted of an Input Layer with 8 input variables + 5 hidden layers and 5 neurons respectively + output layer of 2 classification variables. The accuracy parameter was used as the metric to determine the best-fit model score (77.27 %)

Keras Classifier and K10 fold validation techniques were applied to evaluate the model scores.

1. **Model Evaluation:**

Based on cross-validation, the mean score was 64.97% with a standard deviation of 5.05% from 200 validations. This renders the model moderately well.

A screenshot of a computer program

Description automatically generated

Hence the chosen hidden layers can be considered for building the model.

The neural network with two layers is said to start with input 8 neurons, two hidden layers (5,5) and two output layers.

**Reports in these results:**

The difference between the 5.05% training accuracy and the 64.97% evaluation accuracy is small. This suggests that the model isn't just memorizing the training data but can also work well with new, unseen data.

There are a couple of reasons why the accuracy levels are close. One is that the dataset is big enough, so the model can learn its patterns without just memorizing everything. Another reason could be that the way we set up the model and its settings fit well with the dataset, helping it work with new data too.

But even though the difference is small, there's still a bit of overfitting happening because the training accuracy is lower than the evaluation accuracy. To deal with this, we could try things like dropout, weight decay, or stopping the training early. Also, we might potentially experiment with other model topologies or hyperparameters.

**Report on comparing the accuracy levels from evaluation process:**

The MLP model's accuracy on **the pima\_diabetes\_dataframe** dataset was 64.97%. But let's look at some general things that might affect how well the model works on different datasets. First, bigger, and more complex datasets can be harder for MLP models to learn from. The **Iris** dataset has only four things to look at and 150 examples, while the Pima diabetes one has eight things and 768 examples. This might make it harder for the MLP to do well on the Pima dataset.

**Second**, the accuracy of the model can be affected by how good and relevant the information going into it is. If the input data is messy, repeats itself, or doesn't relate to what we're trying to find out, the model might not do well. So, it's important to clean up and choose the right data for the model to use.

**Third,** how we design the model and pick its settings can also change how accurate it is. Things like how many layers it has, how many parts each layer has, and what math it does at each part can all make a difference. Plus, how fast it learns and how much it learns at once can also matter. So, we need to try out different setups to find the best one for each dataset.

Overall, the accuracy level from the MLP assessment process on **pima\_diabetes\_dataframe** (64.97%) may be impacted by the dataset's complexity and size, the quality and relevance of the input features, and the model architecture and hyperparameters.

**PART III: Redesign the MLP**

MLP Visualization: 8 input variables + 5 neurons +5 neurons + 2 outputs

That dataset, there are several approaches to rebuild the MLP neural network. Increasing the number of hidden layers and/or nodes, adding regularization, or changing the activation functions are all options.

A diagram of a network structure

Description automatically generated

**Preprocess and load the data.**

We can begin by importing and preparing the Pima Indians pima\_diabetes\_data. Scaling the input features, encoding the target variable if necessary, and separating the data into training and testing sets are all possibilities.

A computer screen shot of a program code

Description automatically generated

A screenshot of a computer program

Description automatically generated

* Here's an example of Jupyter not books screenshot.

A screenshot of a computer

Description automatically generated

For several reasons, the updated MLP network could lead to improved performance:

**More Complex Input Data Representation**: By adding more hidden layers and nodes, the MLP network can learn more complex representations of the input data. This allows it to understand more subtle relationships between the input features and the output variable, potentially improving performance.

**ReLU Activation Function:** The Rectified Linear Unit (ReLU) activation function can help mitigate the vanishing gradient problem associated with the sigmoid function. This, in turn, allows the network to train more efficiently and acquire more robust features.

**Dropout Regularization:** Implementing dropouts can effectively reduce overfitting and enhance the network's ability to generalize, enabling it to perform well on unfamiliar data.

When comparing the results obtained from the redesigned MLP with those from part 2, it's important to consider that increasing the number of layers and nodes in a neural network can elevate its complexity and make training more challenging. If the revised MLP wasn't properly tuned or trained for enough epochs, it could potentially perform worse in terms of accuracy or training time compared to the originally planned MLP.

**Results obtained from redesigned MLP, comparing them with those part – 2.**

The gap between the performance of the revised MLP network and the original MLP in part 2 can be attributed to several critical factors:

* Complexity of the Model:
* Training and Hyperparameter Tuning:
* Overfitting:
* Data Characteristics:
* Random Initialization:
* Training Time
* Human Expertise:
* Data Size

Overall, the revised network with three hidden layers and 32 nodes could capture more complicated patterns in data, reduce overfitting, and optimize the learning process, potentially resulting in enhanced performance and higher accuracy levels. However, it is crucial to remember that the model's efficacy is also dependent on the unique qualities of the data and the job at hand, and more testing and tuning may be required to attain the best results.