The Era of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) for Healthcare

Moloud Abdar
Deakin University - Australia
6 April 2021
Emails: mabdar@deakin.edu.au
& m.abdar1987@gmail.com
Outline

- Artificial intelligence, Machine Learning and Deep Learning
- Big Data and AI/ML/DL
- Types of Machine Learning algorithms
- Performance vs Sample Size
- Different Approaches to Solving Problems
- What Exactly Is Deep Learning?
- Why GPU Matters in Deep Learning?
- Sentiment Applications
- Medical Applications of AI/ML/DL
- Detection of Skin Cancer: Deep and Machine Learning Era
- Wart Disease Classification
- Skin Cancer Classification
- Coronary Artery Disease (CAD) Classification
- Detection of Presymptomatic Conditions in Spinocerebellar Ataxia Type 2
Artificial intelligence, Machine Learning and Deep Learning

- **Artificial Intelligence (AI)** is the broader concept of machines being able to carry out tasks in a way that we would consider “smart”.

- **Artificial Intelligences** – devices designed to act intelligently – are often classified into one of two fundamental groups – applied or general.

- **Machine Learning** is a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.

- **Deep learning** structures algorithms in layers to create an "artificial neural network" that can learn and make intelligent decisions on its own.

https://medium.com/@harish_6956/what-is-machine-learning-deep-learning-7788604004da
https://towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac
Big Data and AI/ML/DL

The rise of digital technologies in the last couple of decades allowed information to flourish and grow exponentially.

People and businesses produce huge volumes of data on a daily basis, so it is getting more difficult to analyze information and draw meaningful conclusions out of it.

This is where machine learning and big data come into play.

As such, machine learning helps companies or organizations complete work that people cannot manage single-handedly.

On the other side, big data represents any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.

Types of Machine Learning algorithms

**Supervised learning:** Supervised learning involves training the model on the labeled data and uses this trained model to make predictions on the new data.

**Unsupervised learning:** Unsupervised Learning also involves training of the data except for the fact that the labeled value or target value is not known.

**Semi-supervised learning:** As supervised learning works on labeled data and unsupervised learning on unlabeled data, then a lot of information is lost from labeled data which can be obtained from unlabeled data.

**Reinforcement learning:** Reinforcement Learning works by developing a system which improves its performance by taking feedback from the environment and taking possible steps to improve them.

Performance vs Sample Size

- Almost, no one can tell you how much data you need for your predictive modeling problem.
- It is unknowable: an intractable problem that you must discover answers to through empirical investigation.
- The amount of data required for machine learning depends on many factors, such as:
  - **The complexity of the problem**, nominally the unknown underlying function that best relates your input variables to the output variable.
  - **The complexity of the learning algorithm**, nominally the algorithm used to inductively learn the unknown underlying mapping function from specific examples.

Detection task is harder than classification, but both are almost done.

And with better-than-human quality. (https://research.facebook.com/blog/learning-to-segment/)

Deep learning has an inbuilt automatic multi-stage feature learning process that learns rich hierarchical representations (i.e., features).

What Exactly Is Deep Learning?

• Why is it generally better than other methods on image, speech and certain other types of data?

The short answers:

1. ‘Deep Learning’ means using a neural network with several layers of nodes between input and output,

2. the series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

Figure: https://www.pnas.org/content/116/4/1074
Why GPU Matters in Deep Learning?

- Every set of weights can be stored as a matrix \((m,n)\)

- GPUs are made to do common parallel problems fast. All similar calculations are done at the same time. This extremely boosts the performance in parallel computations.

With GPU, the running time is \(733/27=27.1\) times faster than the running time without GPU!!!
Sentiment Analysis

**Sentiment analysis** (SA, also known as opinion mining or emotion AI) refers to:

I. Natural language processing (NLP),
II. Text analysis,
III. Computational linguistics, and
IV. Biometrics

We use SA to systematically extract, identify, quantify, and study affective states and subjective information.

**SA** is often used in **business** to detect **sentiment** in social data, gauge brand reputation, and understand customers.

Figure: [https://ecommercefastlane.com/how-to-leverage-your-customers-experiences-with-amazon-and-shopify/](https://ecommercefastlane.com/how-to-leverage-your-customers-experiences-with-amazon-and-shopify/)
Visual Question Answering (VQA)

- **VQA** is a recent problem in **computer vision** and Natural language processing (NLP).

- A multi-discipline research problem.

- An algorithm which takes as **input** an **image** and a natural language question about the **image** and generates a **natural language answer** as the **output**.

- The first significant VQA dataset was the **DAtaset for QUestion Answering on Real-world images** (DAQUAR): 6794 training and 5674 test question-answer pairs.

Fake News Detection

- **Fake news** and hoaxes have been there since before the advent of the Internet.

- As social media becomes more and more popular, people tend to consume news more often from it than from traditional news media.

- Fictitious articles deliberately fabricated to deceive readers.

- The fast spreading of fake news stories on social media can cause inestimable social harm.

Figure: [https://www.sciencedirect.com/science/article/pii/S0306457318306794](https://www.sciencedirect.com/science/article/pii/S0306457318306794)
Multi-Modal Fake News Detection

• First study to propose fake news detection for new and time-critical events, which can identify fake news based on multi-modal features and learn transferable features by removing the event-specific features.

• They proposed an end-to-end event adversarial neural networks called EANN.

• The EANN model uses event discriminator to measure the dissimilarities among different events, and further learns the event invariant features which can generalize well for the newly emerged events.

• EANN model is a general framework for fake news detection.

Multi-Modal Fake News Detection

Algorithm 1 Event Adversarial Neural Networks.

Input: The multi-modal input \( \{m_i\}_{i=1}^N \), the auxiliary event label \( \{e_i\}_{i=1}^N \), the detection label \( \{y_i\}_{i=1}^N \) and the learning rate \( \eta \)

1: for number of training iterations do
2:   Decay learning rate according to Eq. 15
3:   Update the parameters of multi-modal feature extractor \( \theta_f \) according to Eq. 14;
4:   Update the parameters of the event discriminator \( \theta_e \):
5:       \[ \theta_e \leftarrow \theta_e - \eta \frac{\partial L_e}{\partial \theta_e} \]
6:   Update the parameters of fake news detector \( \theta_d \):
7:       \[ \theta_d \leftarrow \theta_d - \eta \frac{\partial L_d}{\partial \theta_d} \]
8: end for

Multi-Modal Fake News Detection - Results

Table 1: The Statistics of the Real-World Datasets.

<table>
<thead>
<tr>
<th>Method</th>
<th>Twitter</th>
<th>Weibo</th>
</tr>
</thead>
<tbody>
<tr>
<td># of fake News</td>
<td>7898</td>
<td>4749</td>
</tr>
<tr>
<td># of real News</td>
<td>6026</td>
<td>4779</td>
</tr>
<tr>
<td># of images</td>
<td>514</td>
<td>9528</td>
</tr>
</tbody>
</table>

Table 2: The results of different methods on two datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Vis</td>
<td>0.532</td>
<td>0.598</td>
<td>0.541</td>
<td>0.568</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.596</td>
<td>0.695</td>
<td>0.518</td>
<td>0.593</td>
</tr>
<tr>
<td>Twitter</td>
<td>VQA</td>
<td>0.631</td>
<td>0.765</td>
<td>0.509</td>
<td>0.611</td>
</tr>
<tr>
<td></td>
<td>NeuralTalk</td>
<td>0.610</td>
<td>0.728</td>
<td>0.504</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>att-RNN</td>
<td>0.664</td>
<td>0.749</td>
<td>0.615</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>EANN–EANN</td>
<td>0.648</td>
<td>0.810</td>
<td>0.498</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.715</td>
<td>0.822</td>
<td>0.638</td>
<td>0.719</td>
</tr>
<tr>
<td>Weibo</td>
<td>VQA</td>
<td>0.773</td>
<td>0.780</td>
<td>0.782</td>
<td>0.781</td>
</tr>
<tr>
<td></td>
<td>NeuralTalk</td>
<td>0.717</td>
<td>0.683</td>
<td>0.843</td>
<td>0.754</td>
</tr>
<tr>
<td></td>
<td>att-RNN</td>
<td>0.779</td>
<td>0.778</td>
<td>0.799</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td>EANN–EANN</td>
<td>0.795</td>
<td>0.806</td>
<td>0.795</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.827</td>
<td>0.847</td>
<td>0.812</td>
<td>0.829</td>
</tr>
</tbody>
</table>

Figure 3: The performance comparison for the models w and w/o adversary.

• Using **SA**, you can truly understand the public conversation.

• **Twitter sentiment analysis** allows you to listen to your customers and understand what they need.

• **Sentiment analysis** in *twitter* data using data analytic techniques for predictive modelling.

Energy Choices in Alaska: Geotagged Tweets

- A lexicon-based method for sentiment analysis of tweets is implemented in this study.
- An evidential fuzzy method is proposed to fuse phrase-level sentiment scores into tweet-level overall sentiment scores.
- The rank of top 20 renewable energy-related keywords shows the word “Tidal” has the highest ranking followed by “solar panel”.

Fig. 8. The word cloud of the year (a) 2014, (b) 2015, and (c) 2016 showing the frequency of energy-related keywords in Alaskan tweets.

Fig. 6. The word cloud of the year (a) 2014, (b) 2015, and (c) 2016 showing the frequency of renewable energy-related keywords in Alaskan tweets.

Sentiment Classification of Drug Reviews

- Drug review analysis using 2 deep fusion models based on three-way decision theory.
- First fusion model comprises 3-way fusion of one deep model with traditional method.
- Other model comprises 3-way fusion of three deep models with traditional method.
- Applied 3-way fusion of three deep models with a traditional model (3W3DT) and 3-way fusion of one deep model with a traditional learning algorithm (3W1DT).
- The 3W3DT outperformed 3W1DT by 2% in terms of accuracy and F1-measure.

Medical Applications of AI/ML/DL

Sources:
https://my.vanderbilt.edu/masi/category/big-data/
https://towardsdatascience.com/understanding-cancer-using-machine-learning-84087258ee18
In recent years, healthcare data analysis is becoming one of the most promising research areas.

Big Data is essential to every significant healthcare undertaking. Read about the challenges, applications, and potential brilliant future for healthcare big data.

Using big data analysis to deliver information that is evidence-based will, over time, increase efficiencies and help sharpen our understanding of the best practices associated with any disease, injury or illness.

Healthcare includes data in various types such as Medical image data and Clinical data, Omics data and Sensor data.
Medical Applications of AI/ML/DL

**Medical image** data acquired for different purposes, such as diagnosis, therapy planning, intraoperative navigation, post-operative monitoring and biomedical research.

**Clinical** data includes electronic health records which store patient records collected during ongoing treatment.

**Omics** data is one of the high dimensional data comprising genome, transcriptome and proteome data types.

**Omics** are novel, comprehensive approaches for analysis of complete genetic or molecular profiles of humans and other organisms.

**Sensor** data is collected from various wearable and wireless sensor devices.
Detection of Skin Cancer: Deep and Machine Learning Era

Now, suppose we have some data which were not seen during training of the models:

Note: P is the prediction of the model
Skin Cancer Classification

• Detection of Benign and Malignant skin classes using skin cancer dataset

• Dataset: 2637 pictures (224x244) of the two types of moles for training and 660 pictures (224x244) of the two types of moles for testing.

• Classification method:
  • Application of several deep learning methods (convolutional neural networks)
  • Application of novel Three-way decision (TWD) theory based on entropy
Motivation Behind of Research

- Uncertainty in many real medical tasks:
  - Medical image classification
  - Medical image segmentation
  - Medical signal analysis
  - Computer aided decision support systems for medical applications

- A note about UQ:
  - Alone cannot solve issues.
  - Planning to develop systems based on the uncertainty.

Three-way decisions (TWD)-based BDL (general view)

- Proposed a two-step decision making process using TWD,
- In each level, we use an ensemble Monte Carlo dropout (EMC),
- Decisions based on entropy and STD,
- Avoiding predictions when entropy is less than certainty threshold,
- If model is NOT certain, returns “I DO NOT KNOW”.

The Backbone of DNN Methods

- Applied 4 different well-known DNN methods,

- Add one Flatten layer,

- Two FC layers,

- Two dropouts having 0.5 rate,

- SoftMax layer,

- Tested on Skin cancer dataset

Grad-CAM Visualization

Obtained Results using TWD Theory

<table>
<thead>
<tr>
<th>Method</th>
<th>EC</th>
<th>EI</th>
<th>STD-Co</th>
<th>STD-In</th>
<th>Accuracy (%)</th>
<th>F1 score (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWDBDL (DE)</td>
<td>0.6207</td>
<td>0.6335</td>
<td>0.1351</td>
<td>0.2377</td>
<td>87.55</td>
<td>90.19</td>
<td>0.9377</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWDBDL (EMC)</td>
<td>2.860</td>
<td>3.995</td>
<td>0.2677</td>
<td>0.4724</td>
<td>88.95</td>
<td>89.00</td>
<td>0.920</td>
</tr>
</tbody>
</table>

---

**Phase 1**

**Phase 2**

Wart Disease Classification

- Wart disease (WD) is a skin illness on the human body which is caused by the human papillomavirus (HPV).
- Common warts are small, grainy skin growths that occur most often on your fingers or hands.
- Black dots in the wart are blood vessels that can lead to bleeding.
- Most warts disappear in 1 to 5 years without medical treatment, but treatment is available for warts that are large, numerous, or in sensitive areas.
- Proposed a modified version of Particle swarm optimization (PSO) for optimizing the Artificial Immune Recognition System (AIRS) algorithm called: IAPSO-AIRS.
Material (original data)

The original data set belonged to two different detection methods which were selected from UCI repository dataset as a public free dataset repository [1]:


---

Table 1
Features utilized in the cryotherapy method.

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Values</th>
<th>Mean ± SD³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to treatment</td>
<td>Yes or No</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>47 Man</td>
<td></td>
</tr>
<tr>
<td></td>
<td>43 Woman</td>
<td></td>
</tr>
<tr>
<td>Age (year)</td>
<td>15–67</td>
<td>28.6 ± 13.36</td>
</tr>
<tr>
<td>Time elapsed before treatment (month)</td>
<td>0–12</td>
<td>7.66 ± 3.4</td>
</tr>
<tr>
<td>The number of warts</td>
<td>1–12</td>
<td>5.51 ± 3.57</td>
</tr>
<tr>
<td>Types of wart (Count)</td>
<td>1- Common (54), 2- Plantar (9), 3- Both (27)³</td>
<td></td>
</tr>
<tr>
<td>Surface area of the warts (mm²)</td>
<td>4–750</td>
<td>85.83 ± 131.73</td>
</tr>
</tbody>
</table>

Table 2
Features employed in the immunotherapy method.

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Values</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to treatment</td>
<td>Yes or No</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>41 Man</td>
<td></td>
</tr>
<tr>
<td></td>
<td>49 Woman</td>
<td></td>
</tr>
<tr>
<td>Age (year)</td>
<td>15–56</td>
<td>31.04 ± 12.23</td>
</tr>
<tr>
<td>Time elapsed before treatment (month)</td>
<td>0–12</td>
<td>7.23 ± 3.10</td>
</tr>
<tr>
<td>The number of warts</td>
<td>1–19</td>
<td>6.14 ± 4.2</td>
</tr>
<tr>
<td>Types of wart (Count)</td>
<td>1- Common (47), 2- Plantar (22), 3- Both (21)⁴</td>
<td></td>
</tr>
<tr>
<td>Surface area of the warts (mm²)</td>
<td>6–900</td>
<td>95.7 ± 136.61</td>
</tr>
<tr>
<td>Induration diameter of initial test (mm)</td>
<td>5–70</td>
<td>14.33 ± 17.22</td>
</tr>
</tbody>
</table>

---

³ Standard deviation.

⁴ Patients have both types of common and plantar warts.

⁵ Surface area of biggest wart.

---

We combined both data sets to make a new data set. In addition, we would argue that these treatment methods can be considered as a prediction feature for WD (including 180 records).

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>Range</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Response to treatment</td>
<td>Yes or No</td>
<td>Output (target)</td>
</tr>
<tr>
<td>2.</td>
<td>Gender</td>
<td>88 Men and 92 Women</td>
<td>Input</td>
</tr>
<tr>
<td>3.</td>
<td>Age (years)</td>
<td>15–67</td>
<td>Input</td>
</tr>
<tr>
<td>4.</td>
<td>Time elapsed before treatment (month)</td>
<td>0–12</td>
<td>Input</td>
</tr>
<tr>
<td>5.</td>
<td>The number of warts</td>
<td>1–19</td>
<td>Input</td>
</tr>
<tr>
<td>6.</td>
<td>Types of wart (Count)</td>
<td>1– Common (101), Plantar (31), Both (48)</td>
<td>Input</td>
</tr>
<tr>
<td>7.</td>
<td>Surface area of the warts c (mm²)</td>
<td>4-900</td>
<td>Input</td>
</tr>
<tr>
<td>8.</td>
<td>Procedure of treatment</td>
<td>1=Immunotherapy and 2=Cryotherapy</td>
<td>Input</td>
</tr>
</tbody>
</table>

Results of the Proposed IAPSO-AIRS

Coronary Artery Disease (CAD) Classification

- Coronary artery disease is caused by plaque buildup in the wall of the arteries that supply blood to the heart (called coronary arteries).

- **Risk Factors** for Coronary Artery Disease
  - Physical inactivity
  - Overweight/Obesity
  - Hypertension
  - Smoking
  - Unhealthy diet or eating patterns
  - High cholesterol levels
  - Poorly controlled diabetes
  - Family history of heart disease, especially if a close relative has a history of a cardiac event at an early age, <65yrs.


https://www.uchicagomedicine.org/conditions-services/heart-vascular/coronary-artery-disease
Methodology (the global view)

**Algorithm 1 A General Nested Ensemble (NE) Model**

**Input:** CAD dataset: $D = D_{train} \cup D_{test}$

**Output:** $Q_f$: Best classification output

1. **Begin**
   1. Calculate the portance rate of each feature using a feature selection algorithm;
   2. if $D$ is not balanced then
      1. Balance $D$ using a balancing technique;
   3. end
   4. Select the number of $L$ levels in the NE model;
   5. for $l = 1, \cdots, L$ do
      1. Train the base machine learning algorithms, ensemble learning techniques and metaClassifier at different levels of the NE model using $D_{train}$;
   6. end
   7. Classify unseen records from $D_{test}$ using the NE model;
   8. return Best classification result found;
   9. End

**FIGURE 1.** General scheme of a nested ensemble (NE) model.

The Proposed NE-nu-SVC (the local view)

- NE-nu-SVC: three levels,
- Several ensemble learning
- Multi-step balancing approach
- Genetic-based feature selections algorithm
- Different traditional machine learning methods

Results of the Proposed NE-nu-SVC

<table>
<thead>
<tr>
<th>Measures</th>
<th>Linear</th>
<th>Polynomial</th>
<th>RBF</th>
<th>Sigmoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>37.60</td>
<td>31.89</td>
<td>43.50</td>
<td>30.60</td>
</tr>
<tr>
<td>Normal (%)</td>
<td>5.00</td>
<td>7.00</td>
<td>5.90</td>
<td>7.60</td>
</tr>
<tr>
<td>Average (%)</td>
<td>22.10</td>
<td>25.00</td>
<td>22.60</td>
<td>20.30</td>
</tr>
<tr>
<td>Precision</td>
<td>91.00</td>
<td>89.40</td>
<td>93.30</td>
<td>87.70</td>
</tr>
<tr>
<td>CAD (%)</td>
<td>81.60</td>
<td>82.70</td>
<td>82.30</td>
<td>85.90</td>
</tr>
<tr>
<td>Normal (%)</td>
<td>74.00</td>
<td>77.40</td>
<td>71.70</td>
<td>57.30</td>
</tr>
<tr>
<td>Average (%)</td>
<td>77.50</td>
<td>80.30</td>
<td>77.90</td>
<td>68.90</td>
</tr>
<tr>
<td>Recall</td>
<td>94.60</td>
<td>92.20</td>
<td>91.10</td>
<td>89.40</td>
</tr>
<tr>
<td>CAD (%)</td>
<td>78.20</td>
<td>80.20</td>
<td>78.80</td>
<td>49.40</td>
</tr>
<tr>
<td>Normal (%)</td>
<td>78.20</td>
<td>80.20</td>
<td>78.80</td>
<td>49.40</td>
</tr>
<tr>
<td>Average (%)</td>
<td>78.20</td>
<td>80.20</td>
<td>78.80</td>
<td>49.40</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>78.45</td>
<td>80.41</td>
<td>78.12</td>
<td>49.03</td>
</tr>
</tbody>
</table>

Detection of Presymptomatic Conditions in Spinocerebellar Ataxia Type 2

- Brain is divided into two sections as cerebral cortex and cerebellum.
- Cerebellum influences the balancing of the entire body.
- **Ataxia** is a medical condition in which there is a lack of muscle coordination as a result of an injury to the cerebellum.
- **Ataxia** causes various impairments like ambulatory difficulties, swallowing problems, and other neurological complications.
- **Spinocerebellar ataxia** (SCA) is one of a group of genetic disorders characterized by slowly progressive incoordination of gait and is often associated with poor coordination of hands, speech, and eye movements.

Proposed Method: DL with MCDE and DT


Experimental Results

- Due to the stochastic nature of the procedure, to gather statistically consistent data, there are 10 repeated runs on the same data splits for the entire DL approach using MCD with 500 passes.

- Accordingly, 10 slightly different data sets with features are created and finally the DT are built on each of these.

Any Questions?

THANK YOU FOR YOUR ATTENTION ANY QUESTIONS?