

Superintelligent Deep Learning Artificial Neural Networks

Jamilu Auwalu Adamu

Mathematics Programme, 118 National Mathematical Centre, 904105, FCT-Abuja, Nigeria

Correspondence: Mathematics Programme Building, 118 National Mathematical Centre, Small Sheda, Kwali, FCT-Abuja, Nigeria. Tel: +2348038679094. E-mail: whitehorseconsult@yahoo.com

Abstract

Activation Functions are crucial parts of the Deep Learning Artificial Neural Networks. From the Biological point of view, a neuron is just a node with many inputs and one output. A neural network consists of many interconnected neurons. It is a “simple” device that receives data at the input and provides a response. The function of neurons is to process and transmit information; the neuron is the basic unit in the nervous system. Carly Vandergriendt (2018) stated the human brain at birth consists of an estimated 100 billion Neurons. The ability of a machine to mimic human intelligence is called Machine Learning. Deep Learning Artificial Neural Networks was designed to work like a human brain with the aid of arbitrary choice of Non-linear Activation Functions. Currently, there is no rule of thumb on the choice of Activation Functions, “Try out different things and see what combinations lead to the best performance”, however, sincerely; the choice of Activation Functions should not be Trial and error. Jamilu (2019) proposed that Activation Functions shall be emanated from AI-ML-Purified Data Set and its choice shall satisfy Jameel’s ANNAF Stochastic and or Deterministic Criterion. The objectives of this paper are to propose instances where Deep Learning Artificial Neural Networks are SUPERINTELLIGENT. Using Jameel’s ANNAF Stochastic and or Deterministic Criterion, the paper proposed four classes where Deep Learning Artificial Neural Networks are Superintelligent namely; Stochastic Superintelligent, Deterministic Superintelligent, and Stochastic-Deterministic 1st and 2nd Levels Superintelligence. Also, a Normal Probabilistic-Deterministic case was proposed.

Keywords: Stochastic Superintelligent, Deterministic Superintelligent, Neuron, Stochastic Activation Function, Deterministic Activation Function, Probabilistic, Jameel’s ANNAF Criterion

1. Introduction

Casper Hansen (2019) says “*Better optimized neural network; choose the right activation function, and your neural network can perform vastly better*”.

Artist Hans Hoffman wrote:

“The ability to simplify means to eliminate the unnecessary so that the necessary may speak.”

Activation functions are the fundamental pillars for Artificial Neural Network to learn and make sense of something complicated and Non-linear. They introduce non-linear properties to the Network. If we do not apply a non-linear Activation function then the output signal of Artificial Neural Network would simply be a simple linear function. With a linear function, the network will have limitations to learn complex functional mappings from data. Hence it is very important to provide stable and transparent Activation functions to make the today’s Artificial Neural Network more sophisticated and powerful to learn something complex and complicated form data and represent non-linear complex arbitrary functional mappings between inputs and outputs.

A bigger issue for widespread adoption by large organizations, however, is trust. How much do people trust the AI outputs? How much will people accept what AI systems tell them? The immaturity of some of Artificial Neural Networks’ methodologies exposes the system to a model risk. While it is certainly true that modern AI approaches still have many limitations, there are still plenty of industrial applications that could benefit from AI transformation using the existing techniques in the ever-expanding literature.

A neuron is just a node with many inputs and one output. A neural network consists of many interconnected neurons. It is a “simple” device that receives data at the input and provides a response. First, the neural network learns to correlate incoming and out coming signals with each other - this is called learning.

The function of neurons is to process and transmit information; the neuron is the basic unit in the nervous system. In a simple animal nucleus system, it contains like 302 neurons which allow them to take information in from their surroundings and respond to that information.

The human neuron from Biological perspectives was defined by Carly Vandergriendt (2018) as nerve cells, send and receive signals from your brain. However, they have a lot in common with other types of cells; of course, they are structurally and functionally unique. Specialized projections called axons allow neurons to transmit electrical and chemical signals to other cells. Neurons can also receive these signals via rootlike extensions known as dendrites. Carly Vandergriendt stated the human brain at birth consists of an estimated 100 billion Neurons.

Neurons never reproduced or regenerated but they are only replaced once they die. The parts of Neurons vary in size, shape, and structure depending on their role and location. However, nearly all neurons have three essential parts: a cell body, an axon, and dendrites.

There are three basic functions of neuron as follows:

- Receive signals (or information);
- Integrate incoming signals (to determine whether or not the information should be passed along); and,
- Communicate signals to target cells (other neurons or muscles or glands).

These neuronal functions are reflected in the anatomy of the neuron.

Humans have a hundred billion (100, 000, 000, 000) neurons in our body system, most of those centered in the central nervous system in the brain and the spinal cord. The basic unit of our nervous system is the same; it is a neuron. A body communicates with itself both electrically and chemically. Your nervous system, which includes the brain, spinal cord, and all the nerves, will use electrical and chemical messages to allow you to move, smell, see, feel pain and think. The nervous system is an internal communication system; the main cell of the nervous system is the NEURON. The Neuron transmits nerve impulses, which are electrical signals, and then sends chemicals to the next neuron across a very small gap between them.

The typical neuron has the SAME basic features. The cell body contains the nucleus and many organelles. Dendrites pick up signals from their environment or other neurons, a signal will travel from the Dendrites to the Axon which carries the impulse to the axon terminal, the carries the signal to end of the axon called the axon terminal, messages are then sent in a form of chemical to the next cell.

Therefore, from the above introduction, the basic idea behind artificial neural networks is the structure of the human brain which was mathematically modeled to enable machines to mimic human intelligence through the interconnected Neurons which activates using activation functions. This means, the functioning of Neuron depends on Activation Function, the process “Dendrites pick up signals from their environment or other neurons, a signal will travel from the Dendrites to the Axon which carries the impulse to the axon terminal, the carries the signal to end of the axon called the axon terminal, messages are then send in a form of chemical to the next cell” would not achieved if there does exist a non-linear Activation Function (heart of Neuron). Also, it means for the Neuron to work ROBUSTLY, must have a robust Activation Function. Jamilu (2019) stated that a robust Activation Function shall be EMANATED from AI-ML-Purified Data Set and satisfy Jameel’s ANNAF Stochastic and or Deterministic Criterion.

From the available literature, there is no rule of thumb on the choice of Activation Functions, people try out different things and check the combinations' best performance, and however, the choice of Activation Functions should not be Trial and error. The objective of this paper is to propose instances where Deep Learning Artificial Neural Networks can super-intelligently defeats Human Intelligence. With the aid of robust Activation Functions EMANATED from AI-ML-Purified Data Set satisfy Jameel's ANNAF Stochastic and or Deterministic Criterion, the paper proposed four classes where we can have Super-intelligent Deep Learning Artificial Neural Networks, namely; Stochastic Super-intelligent, Deterministic Super-intelligent and Stochastic-Deterministic 1st and 2nd Super-intelligent. Also, a Normal Probabilistic-Deterministic case was proposed.

2. Materials and Methods

2.1 Materials

2.1.1 Basic Definitions

Linear Function, a linear function is a polynomial of degree one or less, including the zero polynomial. A linear function has the following form: $y = f(x) = a + bx$. A linear function has one independent variable and one dependent variable. The independent variable is x and the dependent variable is y . a is the constant term or the y intercept. It is the value of the dependent variable when $x = 0$. b is the coefficient of the independent variable. It is also known as the slope and gives the rate of change of the dependent variable.

Systems of Non-Linear Equations, a system of equations where at least one equation is not linear is called a nonlinear system.

A *Nonlinear System* is a system in which the change of the output is not proportional to the change of the input. Nonlinear dynamical systems, describing changes in variables over time, may appear chaotic, unpredictable, or counterintuitive, contrasting with much simpler linear systems.. Typically, the behavior of a nonlinear system is described by a nonlinear system of equations.

Differentiable Function, a differentiable function of one real variable is a function whose derivative exists at each point in its domain.

2.1.2 Types of neurons

Neurons vary in structure, function, and genetic makeup. Given the sheer number of neurons, there are thousands of different types, much like there are thousands of species of living organisms on Earth. In terms of function, scientists classify neurons into three broad types: sensory, motor, and interneurons.

Sensory neurons

Sensory neurons help humans to perceive taste, smell, hear, see, and feel things around. Sensory neurons are triggered by physical and chemical inputs from your environment. Sound, touch, heat, and light are physical inputs. Smell and taste are chemical inputs.

Motor neurons

Motor neurons play a role in movement, including voluntary and involuntary movements. These neurons allow the brain and spinal cord to communicate with muscles, organs, and glands all over the body. There are two types of motor neurons: lower and upper. Lower motor neurons carry signals from the spinal cord to the smooth muscles and the skeletal muscles. Upper motor neurons carry signals between your brain and spinal cord.

Interneurons

Interneurons are neural intermediaries found in your brain and spinal cord. They're the most common type of neuron. They pass signals from sensory neurons and other interneurons to motor neurons and other interneurons. Often, they form complex circuits that help you to react to external stimuli.

2.1.3 How Neuron works in a Body System

The cell body has a nucleus, structures arises from cell body call Dendrites, a long fiber extending from the cell body is called Axon, some small branches at the end of axon are called Nerves Endings or Axon Terminals. Nerves help us in information passing from eye to brain in the form of electric impulse from part to another part of our body system. The Dendrites of the Neuron receive the information; the Axon passes the information to the next neurons. When the Dendrites the Neurons received the information, they start a chemical reaction that generates an electric impulse, this electric impulse travel from travels from the Dendrite through the Cell Body along the Axon and finally reaches the Nerves Endings.

The Nerves Endings are not directly connected to the other Neuron because there is some gap between one Nerve Ending of Neuron and Dendritic Tip of the next Neuron, this gap is called Gap/Synapse, now if the neurons are not connected then how the electric impulses pass from one neuron to another neuron, at the end of the axon the electrical impulse triggers the release of some chemicals, that means the electric impulses is converted to some chemicals, these chemicals (signals) crosses the gap to reach the Dendrites of the next Neuron then there will create similar electric impulse, when these chemicals reach the Dendrites of the next Neuron, a similar electric impulse will be created, these electrical passes from the Dendrites to the Axon and again is travel through all the axons and finally reaches the nerves endings. Finally, at the end, the Neurons is connected to the Target Cells by Synapse.

2.2 Methods

2.2.1 Biological versus Mathematical Representations of a Single Neuron

- The current set of Activation Functions is combination of Stochastic and Deterministic Functions (please, see https://en.wikipedia.org/wiki/Activation_function).
- The Neuron receives and transmits information and passes it to the next neuron. The information is a function contains linear, non-linear and probabilistic relationships.
- Thus, with (1) and (2) above concludes that a neural networks are system fundamentally made up of both Stochastic and Deterministic Activation functions. This means, the Hidden and Output Layers neurons' Activation Functions are combination of both Stochastic and Deterministic Functions.

2.2.3 The Proposed Basic Ideas

ACTIVATION FUNCTIONS Allows Artificial Neural Network to learn complex functional mappings from DATA and make sense of something complicated and Non-linear to produce meaningful output signal. Thus, should not be a Trial and error or a Black-Box Assumptions.

Biologically, Neuron performs three basic functions, namely: Receive signals (or information); Integrate incoming signals (to determine whether or not the information should be passed along); and, Communicate signals to target cells (other neurons or muscles or glands). However, this cannot be done successfully without action of a Non-linear Function (Brain of Neuron) residing in a Neuron of Human Brain EMANATED from incoming signals (information). This Non-linear Function (Brain of Neuron) residing in a Neuron of Human Brain EMANATED from incoming signals is what we called Activation Function of a Deep Learning Artificial Neural Network.

The Author PROPOSED that the Competent and Eligible Activation Functions for the successful conduct of Artificial Neural Networks are the Activation Functions EMANATED from the AI-ML-Purified Data Set under consideration satisfied AI-ML-Jameel's Stochastic or Deterministic Criterion because of the following SCIENTIFIC FACTS:

- (1) They EMANATED from the referenced AI-ML-Purified Data Set and satisfied AI-ML-Jameel's Stochastic and or Deterministic Criterion;
- (2) They have a very strong (if not perfect) CORRELATION with the referenced AI-ML-Purified Data Set. A link between the Data Set and Activation Functions MUST be strongly established since Artificial Neural network uses past historical data to predict the future of a given task with the aid of machines;
- (3) They relate better to the referenced AI-ML-Purified Data Set then the existing Assumed and Trial and error Activation Functions;
- (4) They indeed describe the distribution of our referenced AI-ML-Purified Data Set, which is a listing or function showing all the possible values (or intervals) of the data and how often they occur;
- (5) They represent real, virtual and un-virtual information about our referenced AI-ML-Purified Data Set;
- (6) They indeed captured the Symmetric, Left Skewed, Right Skewed, Mesokurtic, Leptokurtic, and Platykurtic properties of our referenced AI-ML-Purified Data Set;
- (7) They contain real, virtual and un-virtual information related to Measures of variability (the range, inter-quartile range, and standard deviation) and Measures of Central Tendency (Mean, Mode and Median, Minimum and Maximum) of our referenced AI-ML-Purified Data Set;
- (8) They capture real, virtual and un-virtual information about the Correlation (autocorrelation) among the elements in our referenced AI-ML-Purified Data Set;
- (9) In the case of Bivariate AI-ML-Data Set, they captured real, virtual and un-virtual information about Measures of Association (Covariance and Correlation) of our referenced AI-ML-Purified Data Set;
- (10) They capture real, virtual and un-virtual information whether or not the parameters of our referenced AI-ML-Purified Data Set are constant over time; and,
- (11) They also captured the presence of *outliers in our* referenced AI-ML-Purified Data Set under consideration.

Referenced AI-ML-Purified Data Set means our referenced Artificial Neural Network Data Set shall possess the following QUALITIES:

- a. Accuracy and Precision
- b. Legitimacy and Validity
- c. Reliability and Consistency
- d. Timeliness and Relevance
- e. Completeness and Comprehensiveness
- f. Availability and Accessibility
- g. Granularity and Uniqueness

Therefore, the author proposed that the practice of "TRIAL AND ERROR" choice of Assumed Activation Functions should be abandoned, however, the choice of Activation Functions should follow a "DEFINITE RULES", and thus, the Author proposed "**JAMEEL'S ANNAF STOCHASTIC AND DETERMINISTIC CRITERION (2019)**" as follows:

2.2.4 Proposed Jameel's ANNAF Stochastic Criterion:

ANNAF means Artificial Neural Network Activation Functions.

Under this criterion, we run the goodness of fits test on our referenced **PURIFIED AI-ML-DATA SET** such that:

- i. We accept if the Average of the ranks of Kolmogorov Smirnov, Anderson Darling and Chi-squared is less than or equal to Three (3);
- ii. We must choose the fat-tailed Probability Distribution $f(x)$ follows by our referenced **PURIFIED AI-ML-DATA SET ITSELF** regardless of its Rankings;
- iii. If there is tie, we include both the fat-tailed Probability Distributions in the selection;
- iv. At least Two (2) fat-tailed Probability Distributions must be included in the selection;
- v. We select the most occur Probability Distribution as the qualify candidate in each case of test of goodness of fit on our referenced **PURIFIED AI-ML-DATA SET**;

- vi. **Backward Propagation Axiom:** Having chosen the most qualified fat-tailed Probability Distribution in the final round up selection for instance $f_{final}(x)$, then $f_{final}(x)$ **SHALL BE MONOTONE CONTINUOUSLY DIFFERENTIABLE FUNCTION.** Particularly if $f_{final}(x)$ is **NOT DIFFERENTIABLE** then **DISCARD** it and repeat (i) to (vi) until we have a MONOTONE DIFFERENTIABLE FUNCTION. Note the Approach is from TOP-BOTTOM process;
- vii. **Criterion Enhancement Axiom:** Thode (2012) intensively discussed about the Best Goodness of Fit Tests such as Kolmogorov Smirnov (KS) Test, Anderson-Darling Test, Jarque and Bera (JB) Test, Shapiro Wilk (SW) Test, Cramer-Von Mises Test, Pearson (χ^2 Godness of Fit) Test, Lilliefors Corrected K-S Test, D'Agostino Skewness Test, Anscombe-Glynn Kurtosis Test, D'Agostino-Pearson and Omnibus Test. Let $\{T_1, T_2, \dots, T_n\}$ be the set of such Best Goodness of Fit Tests, $\{x_1, x_2, \dots, x_n\}$ be their **RANKS** respectively then the generality of (i) can be re-expressed (or enhanced) if $\frac{(x_1 + x_2 + \dots + x_n)}{n} \leq a$, where $0 < a \leq n$, $n \in N$ or equivalently, $x_1 + x_2 + \dots + x_n \leq an$; and,
- viii. **Unit Axiom:** let $f_{final}(x)$ be such that it satisfied axioms (i) to (iv) and or (vii). Let $\{r_1, r_2, \dots, r_n\}$ be the ranks of fitness test of $f_{final}(x)$ obtained from the tests $\{T_1, T_2, \dots, T_n\}$ respectively then if $\forall i \in \{1, 2, \dots, n\}$, $r_i = 1$ regardless of other factors. Consequently, if for all fitness test runs, turn out to be the same $f_{final}(x)$ then $f_{final}(x)$ will gives Deep Learning Artificial Neural Networks **SUPER-INTELLIGENT CAPABILITIES.**

2.2.5 Proposed Jameel's Stochastic Lemma:

All the TOP-RANKED Fat-tailed Monotone Continuously Differentiable Stochastic Functions **EMANATED** from referenced AI-ML-Purified Data satisfied Proposed Jameel's Stochastic ANNAF Criterion are EXCELLENT STOCHASTIC ACTIVATION FUNCTIONS to perform well-informed Forward and Backward Propagations of an Artificial Neural Network.

2.2.6 Propose Jameel's ANNAF Deterministic Criterion

ANNAF means Artificial Neural Network Activation Functions.

For a Neural Network that require **DETERMINISTIC ACTIVATION FUNCTIONS** can satisfy the following proposed criterion:

- (i) The function $f(x)$ shall be **EMANATED** from the referenced AI-ML-Purified Data Set. The essence of the function $f(x)$ to be **EMANATED** from the referenced AI-ML-Purified Data is to build an incredible and sophisticated Activation Function(s) that has the **BEST MATCH AND OR TUNE** with the set of referenced AI-ML-Purified Data since neural network is a system made to learn a function from data. The Activation Functions obtained from the referenced AI-ML-Purified Data can be used to build an extra-ordinary Neural Network Artificial Intelligence System that may defeat Human Intelligence if trained well;
- (ii) A curve fitting for Best Fitted Deterministic Function shall be carried out, the function $f(x)$ whose:
 - (a) Rank is Unity (1)

- (b) Fattiness Standard Error is smaller than any other on the list;
- (iii) The function $f(x)$ shall be Nonlinear;
- (vi) The function $f(x)$ shall have a Range;
- (v) The function $f(x)$ shall be Continuously Differentiable;
- (vi) The function $f(x)$ shall be Monotonic;
- (vii) The function $f(x)$ shall be Smooth Function with a Monotonic Derivative;
- (viii) The function $f(x)$ shall Approximates Identity near the Origin.

If these failed Discard the 1st rated function $f(x)$, repeat (1) to (8) until the qualified Deterministic Activation Function is EMANATED from our referenced AI-ML-Purified Data.

NOTE: Deep Learning Artificial Neural Network's Hidden and output Layers consist of at least one, two or more Best fitted Activation Functions **EMANATED** from our AI-Data Set, therefore, the RANK: UNITY (ONE) in (i)-(a) and Fattiness Standard Error (i)-(b) of the criterion means when a function whose Real "**Rank =1**" was chosen and it satisfied (i) to (viii) then the next function on list whose Real "**Rank=2**" will assume "**New Rank=1**" and will be tested to satisfy all the eight (8) axioms until we have the required number of BEST (EXCELLENT) Activation Functions needed to carry out our Deep Learning Artificial Neural Network.

2.2.7 Proposed Jameel's Deterministic Lemma:

All the TOP-RANKED Nonlinear Monotonic Continuously Differentiable Deterministic Functions **EMANATED** from referenced AI-ML-Purified Data satisfied Proposed Jameel's ANNAF Criterion are EXCELLENT DETERMINISTIC ACTIVATION FUNCTIONS to perform well-informed Forward and Backward Propagations of an Artificial Neural Network.

3. Results

From the available literature, one can only use liner activation function at the output layer, however, apart from that using linear activation function in the hidden layers is composing linear function which is indeed another linear function, unless on other special circumstances.

Depending on the output whether linear or non-linear, one can choose an activation function. It is recommended, if the relationships in the DATA are straight-line relationships, we can use linear Activation function and if the relationships in the DATA are non-linear, we can use non-linear Activation function.

3.1 Output Layer

It is expected at the output, all the neurons of the last layer of the hidden layers will pass the information they received to the output layer. Since all the top-rated Activation Functions are used in the Hidden Layers, it means the system is optimized, hence we have two conditions:

- (1) We may like to further optimized the information received from the last layer of the Hidden Layers, then in this situation, we can use 1st rated stochastic and 1st rated deterministic Activation Functions satisfied Jameel's ANNAF Criterion for the last two neurons.
- (2) We may like to work only with the results of the last layers of the hidden layer thereby normalizing the output, in this situation, we consider the Normal Probability Distribution (or Normal) for stochastic and equivalent non-linear deterministic activation function possesses the same Rank with the Normal Probability Distribution Activation Functions satisfied Jameel's ANNAF- Criterion for the last two neurons.

3.2 Proposed Stochastic Superintelligent Artificial Neural Networks

A Deep Learning Neural Networks whose output is Probabilistic is said to be Stochastically Superintelligent if all the Hidden Layers' NEURONS and Output Layer's NEURONS consists of only 1st Rated Probabilistic Activation Function satisfied ANNAF Stochastic Jameel's Criterion as shown below:

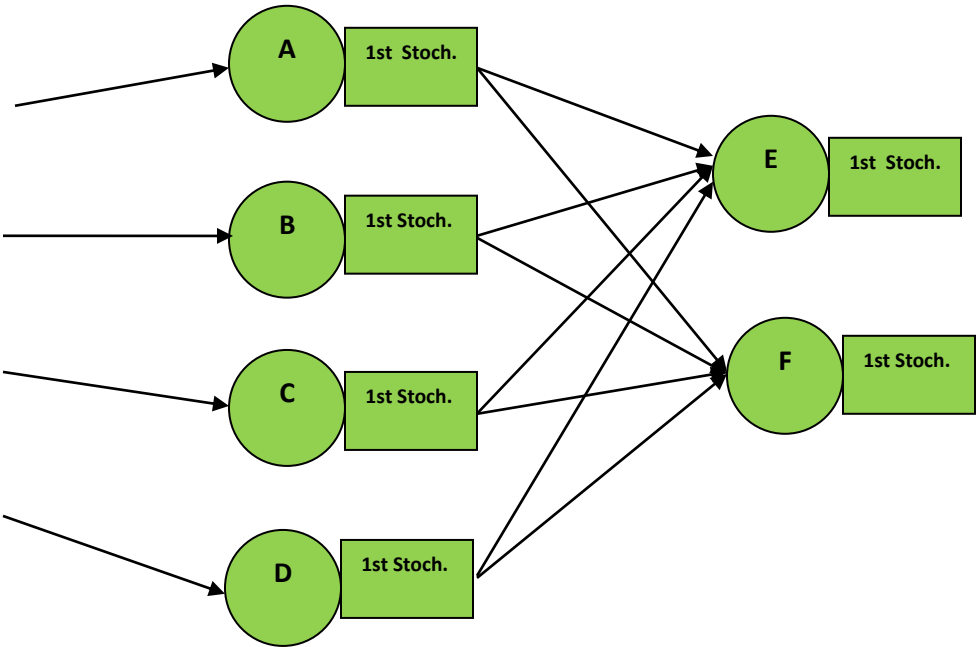


Figure 18: Superintelligent Neural Network of Class A
Source: The Author (2019)

3.3 Proposed Deterministic Superintelligent Artificial Neural Networks

A Deep Learning Neural Networks whose output is Deterministic (Non-probabilistic) is said to be Deterministically Superintelligent if all the Hidden Layers' NEURONS and Output Layer's NEURONS consists of the only 1st Rated Deterministic Activation Function satisfied ANNAF Deterministic Jameel's Criterion as shown below:

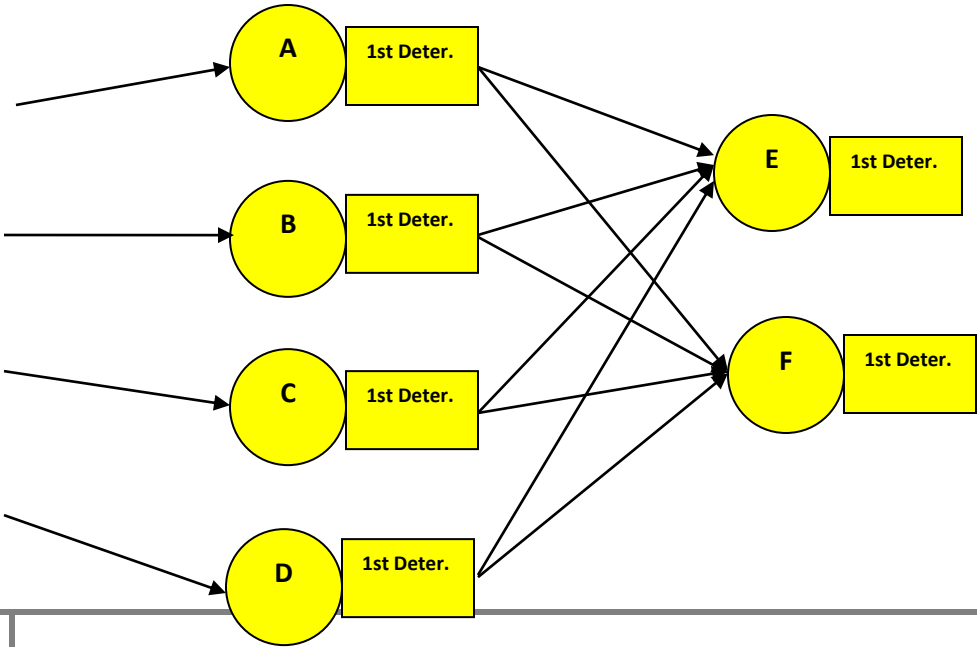


Figure 19: Superintelligent Neural Network of Class B
Source: The Author (2019)

3.4 Proposed 1st Rated Probabilistic-Deterministic Superintelligent Artificial Neural Networks

A Deep Learning Neural Networks whose output is 1st Rated Probabilistic-Deterministic is said to be Stochastically-Deterministically Super-intelligent if all the Hidden Layers' NEURONS and Output Layer's NEURONS are a combination of the only 1st Rated Stochastic and 1st Rated Deterministic Activation Functions satisfied ANNAF Stochastic Jameel's Criterion and AI-ML-Deterministic Jameel's Criterion as shown below:

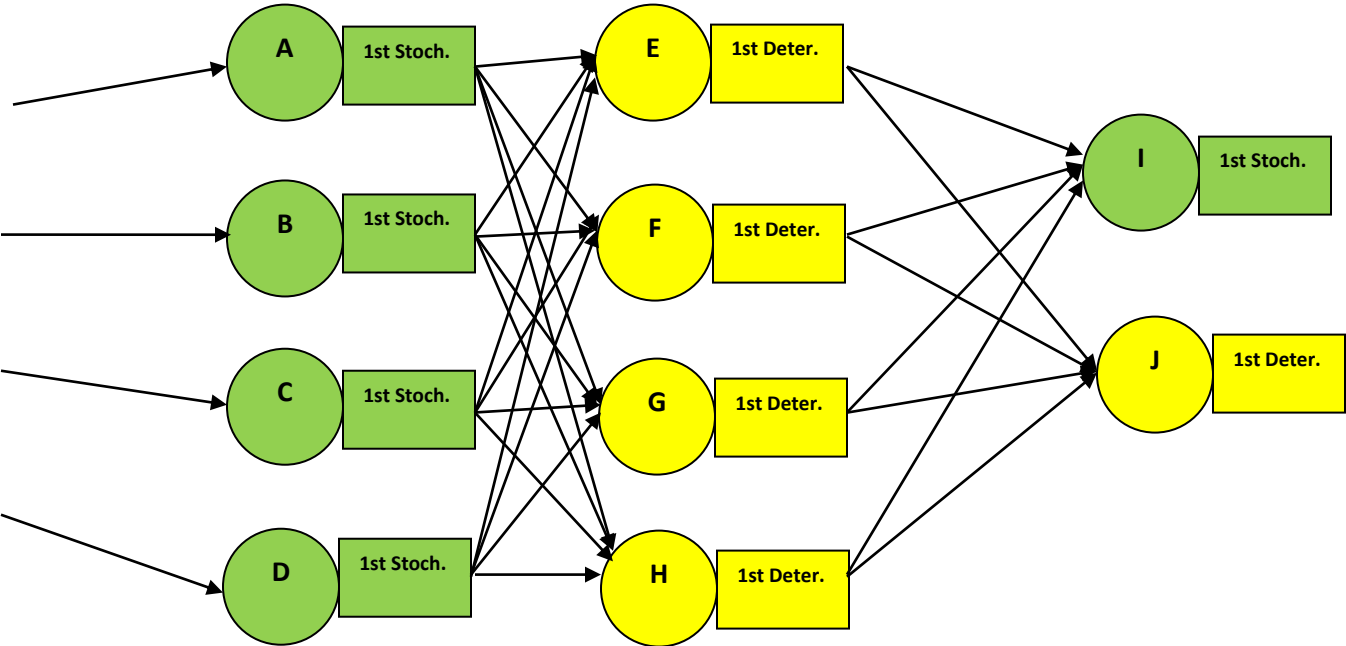


Figure 20: Superintelligent Neural Network of Class C
Source: The Author (2019)

3.5 Proposed 2nd Rated Probabilistic-Deterministic Superintelligent Artificial Neural Networks

A Deep Learning Neural Networks whose output is 1st Rated Probabilistic-Deterministic is said to be 2nd Rated Stochastically-Deterministically Super-intelligent if all the Hidden Layers' NEURONS and Output Layer's NEURONS are a combination of some Top-Rated Stochastic and or Top-Rated Deterministic Activation Functions satisfied AI-ML-Stochastic Jameel's Criterion and ANNAF Deterministic Jameel's Criterion.

Let ATRSODAF represents: Any Activation Function among Top-Rated Stochastic or Deterministic Activation Functions then we have the following Deep Neural Network:

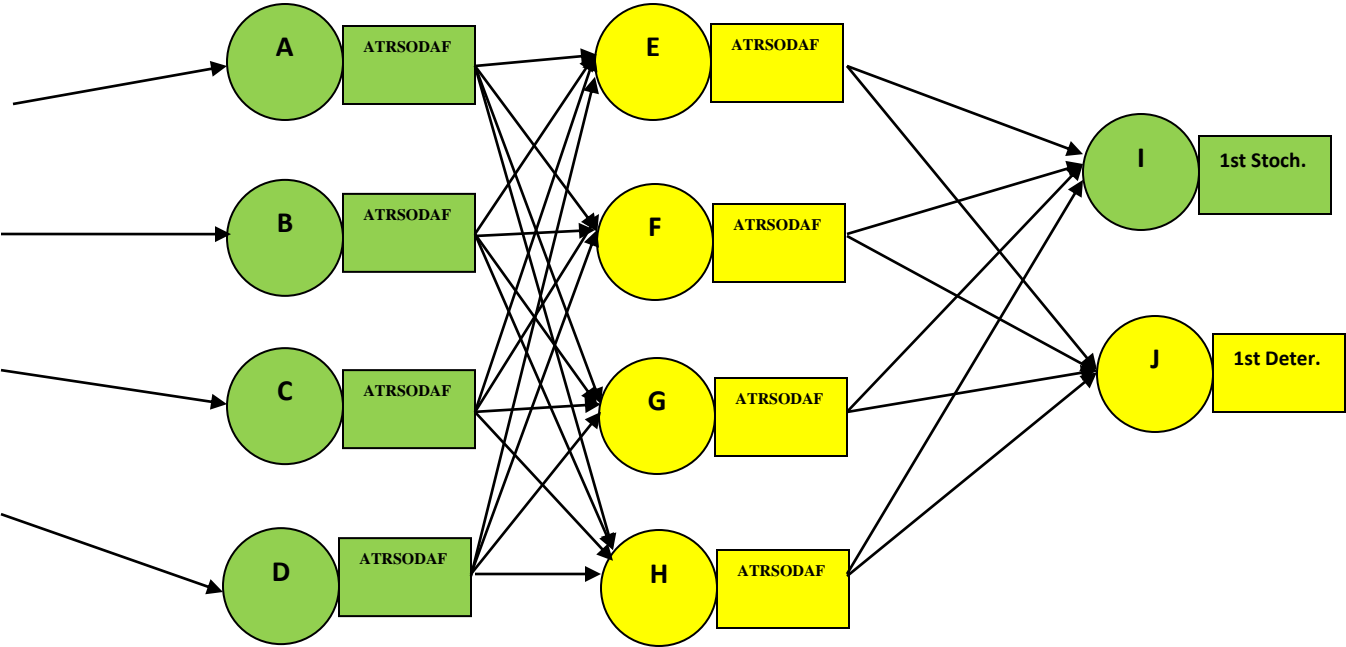


Figure 21: Superintelligent Neural Network of Class D
Source: The Author (2019)
In this case, one may decide to use Linear Activation Function at the output layer if he/she needed linear output.

3.6 Proposed Normal Deep Learning Artificial Neural Networks

A Deep Learning Neural Networks whose output is Normal Probabilistic-Deterministic is said to be Stochastically-Deterministically Normal if all the Hidden Layers’ NEURONS and Output Layer’s NEURONS are combination of only Normal Rated Stochastic and Normal Rated Deterministic (equivalent rating to that of normal stochastic) Activation Functions satisfied ANNAF Stochastic Jameel’s Criterion and ANNAF Deterministic Jameel’s Criterion as shown below:

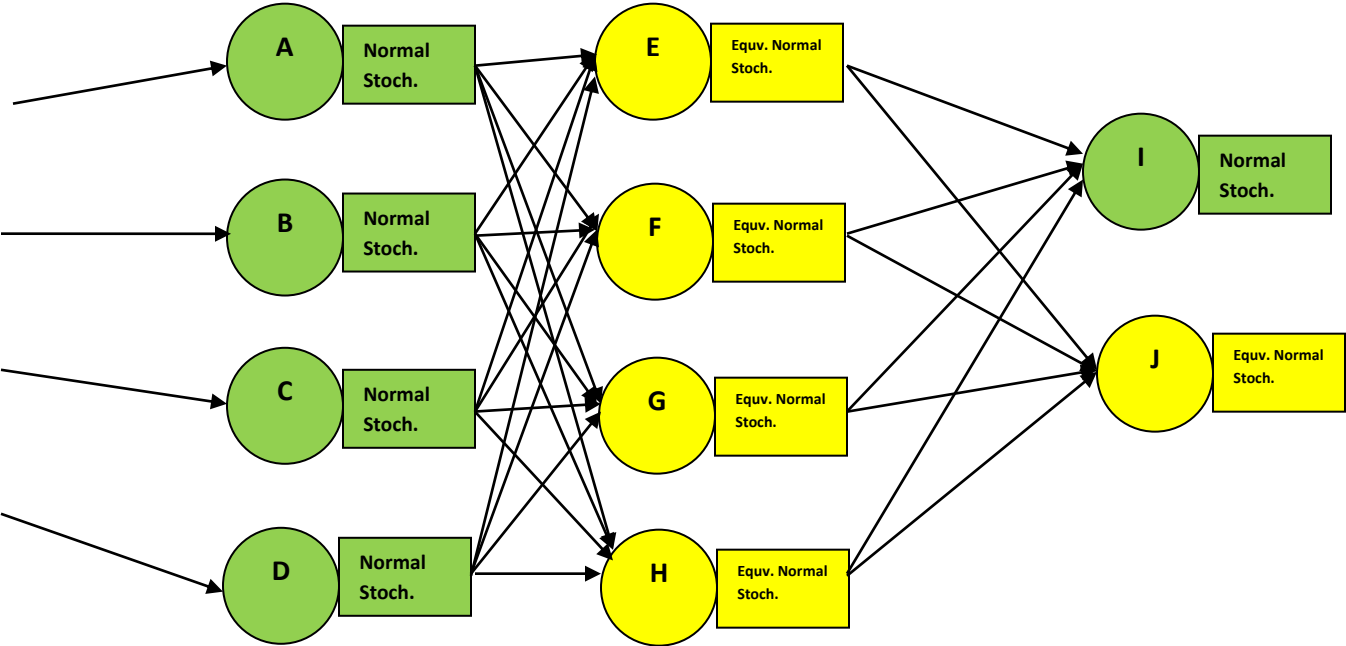


Figure 22: Normal Neural Network of Class E
Source: The Author (2019)

4. Conclusion

The existing Artificial Intelligence Neural Network structure is a system that has no rule of thumb for the choice of its Activation Functions. The hidden layers and output layer's choice of Activation Functions is by trial and error. According to the available literature, one may use ReLu in both the hidden and output layers or Sigmoid in the hidden and output layers or Sigmoid in the hidden layers and Linear at the output layer or Tanh in the hidden layers and Sigmoid at the output layer and so on. Also, it is noted that the existing ANN structure comprises of the arbitrary choice of combinations of both Probabilistic and Deterministic Activation Functions.

This paper proposed with the help of ANNAF Jameel's Stochastic and Deterministic Criterion, four classes of Superintelligent Artificial Neural Networks since their Activation Functions are **EMANATED** from the AI-ML-Purified Data Set and their choices follows a particular Rule of Thumb, not trial and error. This was backed by a Scientific Facts presented in the paper.

The future research direction should be able to test the performances of the proposed four superintelligent classes of Artificial Neural Networks.

Furthermore, this research **REVEALED** that the Advanced Activation Functions satisfied Jameel's ANNAF Stochastic or Deterministic Criterion depends on the **REFERENCED PURIFIED AI DATA SET, TIME CHANGE and AREA OF APPLICATION** (acronym DTA).

This is against the traditional Trial and Error set of assumed Activation Functions, **INDEPENDENT** of **REFERENCED PURIFIED AI DATA SET, TIME CHANGE and AREA OF APPLICATION** (DTA).

This paper was **SUMMARIZED** in the following FIVE (5) Youtube Videos:

- (1) <https://www.youtube.com/watch?v=nth3cJggFts&t=5s>
- (2) <https://www.youtube.com/watch?v=lcyR4TCOBfw>
- (3) <https://www.youtube.com/watch?v=15NgJh71KRQ&t=3s>
- (4) <https://www.youtube.com/watch?v=6emMNluHMZg>
- (5) <https://www.youtube.com/watch?v=IIDTNWc7C-8>

Declaration of Interest: The Author reports no conflict of Interest. The views expressed in this paper are those of the Author and not his current employer.

Acknowledgments

This research paper was a sequel to my Ph.D research work research findings extension with Ahmadu Bello University, Zaria, Nigeria. Firstly, I would like to thank the Federal Government of Nigeria through the National Mathematical Centre for releasing me for my Ph.D study.

My special gratitude and appreciation go to “**TABLECURVE 2D CURVE FITTING**” software of “**SYSTAT**” and “**DERIVATIVE CALCULATOR**”. These materials have tremendously and incredibly increased the speed of the completion of this project work.

Finally, I thank my mum (Hajiya Hauwa Ahmad), my lovely wife Halima and beautiful children Islam and Salman, they are the sources of my Creativity, Energy, and Aspiration.

References

TABLECURVE 2D SOFTWARE, SYSTAT (2019) available online:

<https://systatsoftware.com/products/tablecurve-2d/tablecurve-2d-curve-fitting/>

Derivative Calculator (2019) available on: <https://www.derivative-calculator.net/>

Jamilu Auwalu Adamu (2019), Advanced Deterministic Optimization Algorithm for Deep Learning Artificial Neural Networks, Unpublished manuscript but under review in the Risk and Financial Management Journal

Jamilu Auwalu Adamu (2019), Advanced Stochastic Optimization Algorithm for Deep Learning Artificial Neural Networks in Banking and Finance Industries, Risk and Financial Management Journal, Vol 1 No1 (2019) available online: <https://j.ideasspread.org/index.php/rfm/article/view/387>

Carly Vandergriendt (2018), available online: <https://www.healthline.com/health/neurons>

Nair et al. (2010), Rectified linear units improve restricted boltzmann machines, ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning Pages 807-814, Haifa, Israel — June 21 - 24, 2010.

Djork-Arne Clevert, Thomas Unterthiner & Sepp Hochreiter (2016), FAST AND ACCURATE DEEP NETWORK LEARNING BY EXPONENTIAL LINEAR UNITS (ELUS), Published as a conference paper at ICLR 2016

Klambauer et al. (2017), Self-Normalizing Neural Networks, Institute of Bioinformatics, Johannes Kepler University Linz, Austria.

Lichman, 2013, UCI machine learning repository, URL <http://archive.ics.uci.edu/ml> 901

Aman Dureja and Payal Pahwa (2019), Analysis of Non-Linear Activation Functions for Classification Tasks Using Convolutional Neural Networks, Recent Patents on Computer Science Journal, Volume 12 , Issue 3 , 2019, DOI : [10.2174/2213275911666181025143029](https://doi.org/10.2174/2213275911666181025143029)

Chigozie Enyinna Nwankpa et al. (2018), Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, available online: <https://arxiv.org/pdf/1811.03378.pdf>

Soufiane Hayou et al. (2019), On the Impact of the Activation Function on Deep Neural Networks Training, available online: <https://arxiv.org/pdf/1902.06853.pdf>

Schoenholz et al. (2017), DEEP NEURAL NETWORKS AS GAUSSIAN PROCESSES, Published as a conference paper at ICLR 2018, available online: file:///C:/Users/pc/Downloads/Deep_Neural_Networks_as_Gaussian_Processes.pdf

Asman Dureja and Payal Pahwa (2019), Analysis of Non-Linear Activation Functions for Classification Tasks Using Convolutional Neural Networks, Recent Patents on Computer Science, **Volume 12 , Issue 3 , 2019**, DOI : [10.2174/2213275911666181025143029](https://doi.org/10.2174/2213275911666181025143029)

Casper Hansen (2019) says “Better optimized neural network; choose the right activation function, and your neural network can perform vastly better”, available online: <https://mlfromscratch.com/neural-networks-explained/#/>

Artist Hans Hoffman wrote, “The ability to simplify means to eliminate the unnecessary so that the necessary may speak.” available online: https://www.brainyquote.com/quotes/hans_hofmann_107805

Barnaby Black et.al (2016), Complying with IFRS 9 Impairment Calculations for Retail Portfolios, Moody’s Analytics Risk Perspectives, the convergence of Risk, *Finance, and Accounting*, Volume VII, June, 2016.
Ben Steiner (2019), Model Risk Management for Deep Learning and Alpha Strategies, BNP Paribas Asset Management, *Quant Summit 2019*

Bellotti T. and Crook J. (2012), Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards, *International Journal of Forecasting*, 28(1), 171-182, DOI: 10.1016/j.ijforecast.2010.08.005

Burton G. Malkiel (2009), The Clustering of Extreme Movements: Stock prices and the Weather, *Princeton University*, AtanuSaha, Alixpartners, Alex Grecu, Huron Consulting Group, CEPS working paper No. 186 February, 2009.

Chigozie Enyinna Nwankpa et al. (2018), Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, Preprint.

Daniel Porath (2006), Estimating Probabilities of Default for German Savings Banks and Credit Cooperatives, *University of Applied Sciences, Mainz*, Ander Bruchspitze 50, D – 55122 Mainz

David M. Rowe (2012), Simulating Default Probabilities in Stress Scenarios, Presented to the *PRMIA Global Risk Conference*, New York, NY, May 14, 2012.

David Rich (2019), Responding to the AI Challenge Learning from Physical Industries, @2019 *The Mathworks, Inc.*

Jamilu Auwalu Adamu (2018), Jameel’s Dimensional Stressed Default Probability Models are Indeed IFRS 9 Complaint Models, *Journal of Economics and Management Sciences*, Vol 1 No 2 (2018), pp: 104-114, 2018, DOI: <https://doi.org/10.30560/jems.v1n2p102>.

Jamilu Auwalu Adamu (2017), Jameel’s Criterion and Jameel’s Advanced Stressed Models: An Ideas that Lead to Non-Normal Stocks Brownian Motion Models, *Noble International Journal of Business and Management Research*, Vol. 01, No. 10, pp: 136-154, 2017, URL: <http://napublisher.org/?ic=journals&id=2>.

Jamilu Auwalu Adamu (2016), Reliable and Sophisticated Advanced Stressed Crises Compound Options Pricing Models, *Management and Organizational Studies*, Vol 3, No 1 (2016), pp: 39-55, 2016, DOI: <https://doi.org/10.5430/mos.v3n1p39>.

Jamilu Auwalu Adamu (2017), Jameel’s Criterion and Jameel’s Advanced Stressed Models: An Ideas that Lead to Non-Normal Stocks Brownian Motion Models, *Noble International Journal of Business and Management Research*, Vol. 01, No. 10, pp: 136-154, 2017, URL: <http://napublisher.org/?ic=journals&id=2>

Jamilu Auwalu Adamu (2017), An Introduction of Jameel’s Advanced Stressed Economic and Financial Crises Models and to Dramatically Increasing Markets Confidence and Drastically Decreasing Markets Risks, *International Journal of Social Science Studies*, Vol 4, No 3 (2016), pp: 39-71, DOI: <https://doi.org/10.11114/ijsss.v4i3.1326>

- Jamilu Auwalu Adamu (2015), Banking and Economic Advanced Stressed Probability of Default Models, *Asian Journal of Management Sciences*, 03(08), 2015, 10-18.
- Jamilu A. Adamu (2015), Estimation of Probability of Default using Advanced Stressed Probability of Default Models, Ongoing Ph.D Thesis, Ahmadu Bello University (ABU), Zaria, Nigeria.
- Joonho Lee et al. (2019), ProbAct: A Probabilistic Activation Function for Deep Neural Networks, Preprint. Under review.
- Nassim N. Taleb (2011), The Future has Thicker Tails Past: Model Error as Branching Counterfactuals, presented in Honor of Benoit Mandelbrot's at his Scientific Memorial, Yale University, April, 2011.
- Nassim N. Taleb (2011), A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error, First Version, 2011
- Nassim N. Taleb (2010), Why Did the Crisis of 2008 Happen, Draft, 3rd Version, August, 2010.
- Nassim N. Taleb (2009), Errors, Robustness, and Fourth Quadrant, New York University Polytechnic Institute and Universa Investment, United States, *International Journal of Forecasting* 25 (2009) 744 – 759
- Nassim N. Taleb (2010), Convexity, Robustness, and Model Error inside the “Black Swan Domain”, Draft Version, September, 2010.
- Nassim N. Taleb et al (2009), Risk Externalities and Too bid to Fail, New York University Polytechnic Institute, 11201, New York, United States.
- Nassim N. Taleb (2012), The Illusion of Thin – Tails under Aggregation, NYU – Poly, January, 2012
- Nassim N. Taleb (2007), Black Swans and the Domains of Statistics, *American Statistician*, August 2007, Vol. 61, No. 3.
- Mohit Goyal et al. (2019), Learning Activation Functions: A new paradigm for understanding Neural Networks, *Proceedings of Machine Learning Research* 101:1–18, 2019.
- M & Van Ness (1968), Fractional Brownian Motions, Fractional Noises and Applications (M & Van Ness (1968)), *SIAM Review*: 10, 1968, 422-437
- Onali E. & Ginesti G. (2014), Pre-adoption Market Reaction to IFRS 9: A Cross-country Event-study, *Journal of Accounting and Public Policy*, 33(6), 628-637.
- Peter Martey Addo et al.(2018), Credit Risk Analysis using Machine and Deep Learning Models, *Risks Journal*, Risks 2018,6,38;doi:10.3390/risks6020038.
- Ram Ananth et al. (2019), Opening the “Black Box” , The Path to Deployment of AI Models in Banking, White Paper, DataRobot and REPLY AVANTAGE.
- Reney D. Estember and Michael R. Marana (2016), Forecasting of Stock Prices using Brownian Motion - Monte Carlo Simulation, *Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management*, Kuala Lumpur, Malaysia, March 8-10, 2016.
- Sebastian Urban (2017), Neural Network Architectures and Activation Functions: A Guassian Process Approach, Technical University Munich, 2017.
- Sebastian Raschka (2018), STAT 479: Machine Learning Lecture Notes, <http://stat.wisc.edu/~sraschka/teaching/stat479fs2018/>.
- Sven-Patrik Hallsjo, Machine Learning, Deep Learning, Experimental Particle Physics, University of Glasgow.
- Steven R. Dunbar, Stochastic Processes and Advanced Mathematical Finance, The Definition of Brownian Motion and the Wiener process, Department of Mathematics, 203 Avery Hall, University of Nebraska-Lincoln, Lincoln, NE 68588-0130
- Spreedhar T Bharath et al (2004), Forecasting Default with the KMV – Merton Model, University of Michigan, Ann Arbor MI 48109.
- Soufiane Hayou et al. (2019), On the Impact of the Activation Function on Deep Neural Networks Training, *Proceedings of the 36 th International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019
- TidarukAreerak (2014), Mathematical Model of Stock Prices via a Fractional Brownian Motion Model with Adaptive Parameters
- Ton Dieker (2004), Simulation of Fractional Brownian Motion, Thesis, University of Twente, Department of Mathematical Sciences, P.O. BOX 217, 7500 AE Enschede, Netherlands

Wenyu Zhang (2015), Introduction to Ito's Lemma, Lecture Note, Cornell University, Department of Statistical Sciences, May 6, 2015.

<https://www.stoodnt.com/blog/scopes-of-machine-learning-and-artificial-intelligence-in-banking-financial-services-ml-ai-the-future-of-fintechs/>

<https://medium.com/datadriveninvestor/neural-networks-activation-functions-e371202b56ff>

<https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/>

<http://www.datastuff.tech/machine-learning/why-do-neural-networks-need-an-activation-function/>

<https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>

<https://www.youthkiawaaz.com/2019/07/future-of-artificial-intelligence-in-banks/>

<https://news.efinancialcareers.com/uk-en/328299/ai-in-trading-buy-side>

<https://ai.stackexchange.com/questions/7609/is-nassim-taleb-right-about-ai-not-being-able-to-accurately-predict-certain-type/7610>

<https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c>

The U.S. Hearing Youtube (Technology Companies & Algorithms):

<https://www.youtube.com/watch?v=vtw4e68CkwU>

<https://www.commerce.senate.gov/2019/6/optimizing-for-engagement-understanding-the-use-of-persuasive-technology-on-internet-platforms>

<https://ai.stackexchange.com/questions/7088/how-to-choose-an-activation-function>

<https://mlfromscratch.com/activation-functions-explained/#/>

<https://github.com/nadavo/mood>.

Backward propagation and Activation Functions:

<https://www.youtube.com/watch?v=q555kfIFUCM>

<https://www.youtube.com/watch?v=-7scQpJT7uo>

<https://towardsdatascience.com/analyzing-different-types-of-activation-functions-in-neural-networks-which-one-to-prefer-e11649256209>

<http://vision.stanford.edu/teaching/cs231n-demos/linear-classify/>

<http://vision.stanford.edu/teaching/cs231n-demos/knn/>

<https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle®Dataset=reg-plane&learningRate=0.03®ularizationRate=0&noise=0&networkShape=4,2&seed=0.41357&showTestData=false&discretize=false&percTrainData=50&x=true&y=true&xTimesY=false&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false>

