MACHINE LEARNING WITH APPLICATIONS IN AGRICULTURE

ABIODUN UTHMAN ALLISON
MATRIC NO: 204603

A M.Sc. RESEARCH PROJECT SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITY OF IBADAN, IBADAN, NIGERIA.

SUPERVISOR: PROF. G.O.S. EKHAGUERE

FEBRUARY, 2021

TABLE OF CONTENTS

1 Introduction

2 Literature Review

3 Pricing of Options with Default Risks under The First Passage Time Model
- Reduction to Pricing of Options without Default Risks
- Extension to Structured Note Pricing
- Numerical Results

4 Conclusion

Introduction I

Agriculture could be referred to as the production, processing, promotion and distribution of agricultural products. It plays a critical role in the economy of developing countries. Besides supplying food and raw materials, agriculture also provides the populace with job opportunities. The Food and Agriculture Organization of the United Nations (FAO) estimates that pests and diseases lead to the loss of $20 - 40\%$ of global food production, constituting a threat to food security (Food and Agriculture Organization of the United Nation, International Plant Protection Convention, 2017).
Machine learning is the scientific study of algorithms and mathematical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is a branch of Artificial Intelligence (AI) focused on the premise that, with minimal human interaction, systems can learn from data, recognize trends and make decisions. In supervised machine learning, a training set

Introduction II

of D input vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D\}$ is used to tune the parameters of an adaptive model, along with their corresponding target vectors $\boldsymbol{y}$, in order to accurately predict vector $\hat{\boldsymbol{y}}$. We focus on a subset of supervised learning, called classification, where the aim is to assign each input vector to one of a finite number of discrete categories

Computer vision, and object recognition in particular, has made tremendous advances in the past few years. In 2012, a large, deep convolutional neural network achieved a top-5 test error (the fraction of test images for which the correct label is not among the five labels considered most probable by the model) rate of 15.3 %, at the Large Scale Visual Recognition Challenge (ILSVRC) 2012 image classification competition for the classification of 15 million images into 22 thousand possible categories **Krizhevsky et al. (2012)**.

Literature Review I

Johnson and Stulz, 1987 was the first paper covering default risk in options. They considered many options with writer's default risk which are written by private financial institutions that have limited assets instead of organized exchanges and there are no guarantee by a third party. Therefore, they assumed that if the option writer is unable to make a promised payment, the vulnerable option holders receive all the assets of the option writer. They derived closed form solutions for the European option prices in some different cases by assuming stochastic processes both for the asset value of the option writer and the value of the underlying asset.

Zhou and Li, 2019 considered the pricing problem of options with counterparty default risks. They studied the asymptotic behaviour of vulnerable option prices under an uncertain volatility model which contains both corporate assets and underlying assets. They presented an estimation method for solving a fully non-linear P.D.E by imposing

Literature Review II

additional conditions to the boundary conditions and substituted it into two Black-Scholes like equations whereby they obtained a method to estimate the price of vulnerable options in the worst case scenario.

Theorem I

Assume $B_1(t), B_2(t)$ are correlated Brownian motions with drift terms $\alpha t, \beta t$ and $dB_1(t)dB_2(t) = \rho dt$. Then

$$\int_{\bar{b}_2}^{\infty}\int_{\bar{b}_1}^{\infty}\int_{\bar{m}_1}^{\infty} e^{pb_1+qb_2} f_{M_1(T),B_1(T),B_2(T)}(m_1, b_1, b_2)dm_1 db_1 db_2$$

$$= e^{T(\frac{p^2}{2}+\rho pq+\frac{q^2}{2}+\alpha p+\beta q)}$$

$$N_2\left(-\frac{\bar{b}_1}{\sqrt{T}} + \sqrt{T}(\alpha + p + \rho q), -\frac{\bar{b}_2}{\sqrt{T}} + \sqrt{T}(\beta + \rho p + q), \rho \right)$$

$$- e^{T(\frac{p^2}{2}+\rho pq+\frac{q^2}{2}+\alpha p+\beta q)+2\bar{m}_1(p+\rho q)+2\alpha\bar{m}_1}$$

$$N_2\left(-\frac{\bar{b}_1 - 2\bar{m}_1}{\sqrt{T}} + \sqrt{T}(\alpha + p + \rho q), -\frac{\bar{b}_2 - 2\rho\bar{m}_1}{\sqrt{T}} + \sqrt{T}(\beta + \rho p + q), \rho \right)$$

Theorem II

where $N_2(u, v, \rho)$ is the cumulative distribution function of a bivariate standard normal distribution defined by

$$N_2(u, v, \rho) \equiv \int_{-\infty}^{v} \int_{-\infty}^{u} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)} dxdy$$

Substituting $\bar{b}_1 = d, \bar{b}_2 = k, \bar{m}_1 = l, \alpha = B_1(t), \beta = B_2(t)$, equation (**??**) can be simplified with Theorem 1 as follows:

$$C = (1-\delta)S_0\left[N_2\left(-\frac{d}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2}+\rho\sigma_s\sigma_v),\right.\right. \\ \left.\left.-\frac{k}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_s}(r+\frac{\sigma_s^2}{2}),\rho\right)\right. \tag{1}$$

$$\left. - e^{2\rho\sigma_s l + \frac{2l}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})} N_2\left(\frac{2l-d}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2}+\rho\sigma_s\sigma_v),\right.\right. \\ \left.\left.\frac{2\rho l - k}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_s}(r+\frac{\sigma_s^2}{2}),\rho\right)\right]$$

$$- (1-\delta)Ke^{-rT}\left[N_2\left(-\frac{d}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2}), -\frac{k}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_s}(r-\frac{\sigma_s^2}{2}),\rho\right)\right. \tag{2}$$

$$- e^{\frac{2l}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})} N_2\left(\frac{2l-d}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2}), \frac{2\rho l-k}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_s}(r-\frac{\sigma_s^2}{2}), \rho\right)\Big]$$
$$+ \delta(S_0 N_1(d_1) - Ke^{-rT} N_1(d_2))$$

Recall that

$$d_1 = \frac{1}{\sigma_s\sqrt{T}} \ln\left[\frac{S_0 e^{(r+\frac{\sigma_s^2}{2})T}}{K}\right], \quad d_2 = \frac{1}{\sigma_s\sqrt{T}} \ln\left[\frac{S_0 e^{(r-\frac{\sigma_s^2}{2})T}}{K}\right]$$

Setting;

$$d_3 = \frac{1}{\sigma_v\sqrt{T}} \ln\left[\frac{V_0 e^{(r+\frac{\sigma_v^2}{2})T}}{D}\right], \quad d_4 = \frac{1}{\sigma_v\sqrt{T}} \ln\left[\frac{V_0 e^{(r-\frac{\sigma_v^2}{2})T}}{D}\right]$$

By direct computation, $-\frac{k}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_s}(r+\frac{\sigma_s^2}{2}) = d_1$ and $-\frac{d}{\sqrt{T}} + \frac{\sqrt{T}}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2}) = d_4$

Having obtained all the values, the pricing formula for a European call option with default risk is given by:

$$
\begin{aligned}
C = (1-\delta)S_0 \Bigg[& N_2(d_4 + \sqrt{T}\rho\sigma_s, d_1, \rho) \\
& - \underbrace{e^{2\rho\sigma_s I + \frac{2I}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})} N_2\left(d_4 + \frac{2I}{\sqrt{T}} + \sqrt{T}\rho\sigma_s, d_1 + \frac{2\rho I}{\sqrt{T}}, \rho\right)}_{A} \Bigg] \\
& - (1-\delta)Ke^{-rT}\Bigg[N_2(d_4, d_2, \rho) \\
& - \underbrace{e^{\frac{2I}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})} N_2\left(d_4 + \frac{2I}{\sqrt{T}}, d_2 + \frac{2\rho I}{\sqrt{T}}, \rho\right)}_{B} \Bigg] \\
& + \delta[S_0 N_1(d_1) - Ke^{-rT} N_1(d_2)]
\end{aligned}
\tag{3}
$$

A European put option with default risk can also be evaluated by

Reduction to Pricing of Options without Default Risks I

The value of a European call option with default risk given in equation (3) boils down to the Black-Scholes formula when the default risk of writer is ignored. Since the last term of equation (3) is just δ times of Black-Scholes pricing formula for European call option, we only need to show the rest terms reduce to $(1-\delta)C_{BS}$ to verify the whole reduction.
Setting default boundary $\tilde{L}(t)$ to zero, we have $D \to 0, L \to 0$, where

$$\lim_{d_4 \to \infty} N_2(d_4, d_2, \rho) = \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$
$$= N_1(d_2) \quad \text{by definition}$$

Reduction to Pricing of Options without Default Risks II

Equation (3) leads to the conclusion given by

$$C = S_0 N_1(d_1) - K e^{-rT} N_1(d_2) \tag{5}$$

Extension to Structured Note Pricing I

The mentioned pricing methodology can also be used to evaluate the price of a structured note with default risks. The pricing of a reverse exchangeable bond under the first passage model is shown below.
A reverse exchangeable bond is a structured note, in which the holder will receive returns depending on an underlying asset at maturity date T. Let $S(t)$ be the value process of underlying asset. The holder receives principal and coupon if $S(T) \geq S(0)$. Otherwise, the holder receives principal and underlying asset if $S(T) < S(0)$. The straight bond with default risk's value is obtained as

$$B = e^{-rT}(F+FcT)\left(\delta + (1-\delta)\left(N_1(d_4) - e^{\frac{2l}{\sigma_v}(r-\gamma-\sigma_v^2/2)}N_1\left(d_4 + \frac{2l}{\sqrt{T}}\right)\right)\right)$$

Extension to Structured Note Pricing II

The pricing formula for a reverse exchangeable bond with default risk can be derived as

$$B - \left(\frac{F}{K}\right) P$$

$$= e^{-rT}(F + FcT)\left(\delta + (1-\delta)\left(N_1(d_4) - e^{\frac{2l}{\sigma_v}\left(r-\gamma-\frac{\sigma_v^2}{2}\right)} N_1\left(d_4 + \frac{2l}{\sqrt{T}}\right)\right)\right)$$

$$- \left(\frac{F}{K}\right)\left[(1-\delta)Ke^{-rT}\left[N_2(d_4, -d_2, -\rho) - e^{\frac{2l}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})}\right.\right.$$

$$\left.\left. N_2\left(d_4 + \frac{2l}{\sqrt{T}}, -d_2 - \frac{2\rho l}{\sqrt{T}}, -\rho\right)\right]\right.$$

$$\left(- (1-\delta)S_0\left[N_2(d_4 + \sqrt{T}\rho\sigma_s, -d_1, -\rho) - e^{2\rho\sigma_s l + \frac{2l}{\sigma_v}(r-\gamma-\frac{\sigma_v^2}{2})}\right.\right.$$

$$\left.\left. N_2\left(d_4 + \frac{2l}{\sqrt{T}} + \sqrt{T}\rho\sigma_s, -d_1 - \frac{2\rho l}{\sqrt{T}}, -\rho\right)\right]\right)$$

Numerical Results I

Ignoring the default risk of counterparty causes great danger. In this section, we show how significant an option could be mispriced if the default risk of the counterparty is ignored. With parameters $K = 40, V(0) = 100, S(0) = 40, D = 90, \gamma = 0.06, \delta = 0.6, T = 3, L = 70, \rho = 0, \sigma_v = \sigma_s = 0.2$ and $r = 0.05$.
To continue in our analysis, we will use MATLAB to calculate the pricing formula of a European call option with default risk and European call option without default risk (Black-Scholes formula). Also, using MATLAB, we will show graphically, the comparison of values of a European call option with default risk and one without default risk. The option values under different $S(0), V(0), \sigma_v$ and T will be shown respectively.

Numerical Results II

Let $d_1 = 0.6062, d_2 = 0.2598, d_3 = 0.9104, d_4 = 0.5640, N_2(d_4, d_2, \rho) = 0.4300,$
$I = -2.6834, N_2(d_4 + \frac{2I}{\sqrt{T}} + \sqrt{T}\rho\sigma_s, d_1 + \frac{2\rho I}{\sqrt{T}}, \rho) = 0.0041,$
$N_2(d_4+\sqrt{T}\rho\sigma_s, d_1, \rho) = 0.5194, N_2(d_4+\frac{2I}{\sqrt{T}}, d_2+\frac{2\rho I}{\sqrt{T}}, \rho) = 0.0034,$
$N_1(d_1) = 0.7278, N_1(d_2) = 0.6025, K = 40, V(0) = 100, S(0) = 40,$
$D = 90, \gamma = 0.06, \delta = 0.6, T = 3, L = 70, \rho = 0, \sigma_v = \sigma_s = 0.2$ *and* $r = 0.05$
From equation (3), the price of a European option with default risk is given as

$$C = 7.3695 \tag{6}$$

Numerical Results III

Also, the price of a European call option without default risk is:

$$C = 8.3689 \tag{7}$$

S(0)	20	25	30	35	40	45	50	55	60
CBS	0.2239	0.9699	2.5579	5.0663	8.3689	12.2754	16.5971	21.1830	25.9306
C	0.3955	0.8505	2.2485	4.4603	7.3695	10.8063	14.6135	18.649	22.8289

Table: Values of European call option with default risk and European call option without default risk as S(0) varies

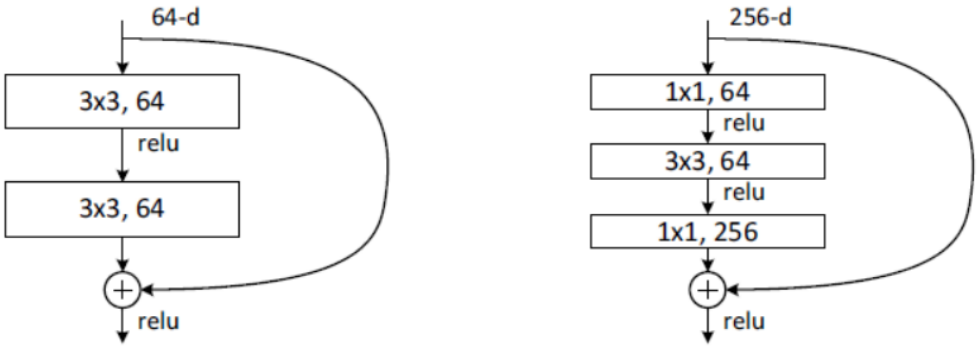

Figure: Comparison of the values of European call option with default risk and European call option without default risk as $S(0)$ varies

The values of European call options are shown in Table 4.1 and Figure 4.1, where $S(0)$ varies from 20 to 60, the price of European call options with default risks are evaluated by equation (3) and the price of European call options without default risk, the Black-Scholes formula. Since the payoff function $(S(T) - K)^+$ has more probability to be large if $S(0)$ is large, the value of a European call option increases as $S(0)$ increases. This can be observed in the case of a European option with default risk, as well as European option without default risk. As shown Figure 4.1, the value of an option with default risk is lower than one without default risk.

V(0)	100	120	140	160	180	200	220	240	260	280	300
CBS	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689
C	7.3694	7.9688	8.1536	8.2768	8.3287	8.3508	8.3614	8.3665	8.3688	8.3677	8.3694

Table: Values of European call option with default risk and European call option without default risk as $V(0)$ varies

The difference between these two derivatives due to the default risk of counterparty can be thought of as a risk premium, which is even more significant when $S(0)$ is fixed and $V(0)$ is changed as shown in Table 4.2 and Figure 4.2.

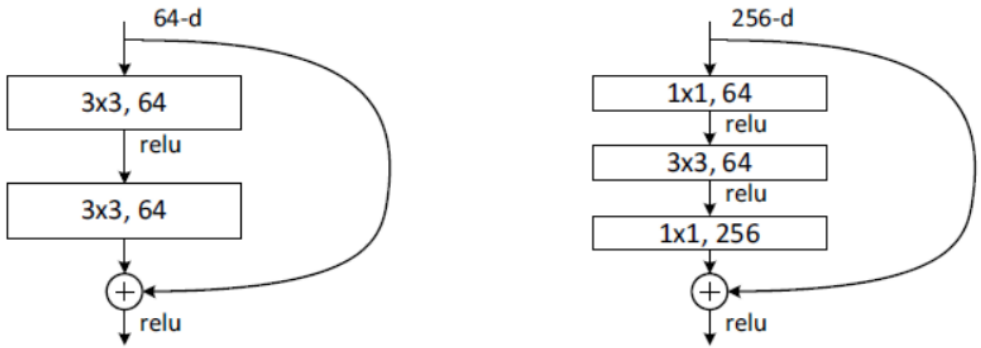

Figure: Comparison of the values of European call option with default risk and European call option without default risk as $V(0)$ varies

σ_v	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
CBS	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689	8.3689
C	0	8.0005	7.3695	6.7902	6.337	6.0072	5.7689	5.5933	5.4637	5.3655	5.2896

Table: Values of European call option with default risk and European call option without default risk as σ_v varies

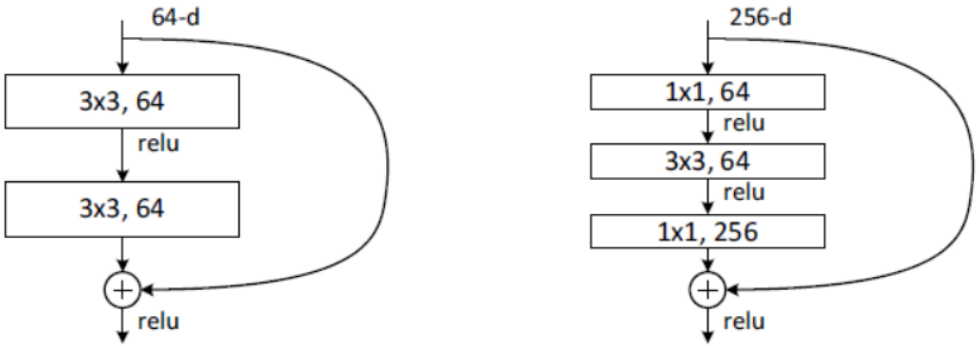

Figure: Comparison of the values of European call option with default risk and European call option without default risk as σ_v varies

Similarly, under the same setting but different σ_v , the values of European call option with default risk and European call option without default risk also have difference which varied significantly. As shown in Table 4.3 and Figure 4.3, when σ_v is large, ignoring the default risk of counterparty and directly using the Black-Scholes formula to assess an option with default risk will cause significant overpricing. This is due to the great default risk of counterparty when its volatility σ_v is large or initial firm value $V(0)$ is too small. In contrast, when $V(0)$ is large or σ_v is too small, the value of a European option with default risk tends to that of European option without default risk.

T	1	1.5	2	2.5	3
CBS	4.1837	5.3759	6.4510	7.4424	8.3689
C	3.766	4.7952	5.7292	6.5797	7.3695

Table: Values of European call option with default risk and European call option without default risk as T varies

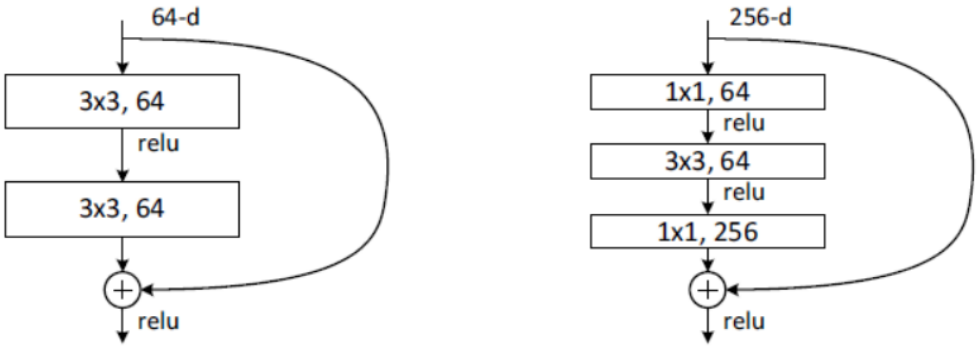

Figure: Comparison of the values of European call option with default risk and European call option without default risk as T varies

Also, Table 4.4 and Figure 4.4 show the values of European call option with default risk and European call option without default risk as maturity date T varies. Here, as T increases, the values of European call option with default risk and European call option without default risk also increase.
The black line denotes the value of European call options with default risks and the red line shows the values of European call options without default risk, evaluated by Black-Scholes formula.

The above figures show that an option holder who buy the option from a small firm, in contrast to a big firm, is more vulnerable and hence needs more risk premium, which is consistent with the intuition.

Conclusion I

In this project, we derived the analytical pricing formula for European options with default risks, releasing the constrain of default risk model used to measure the default risk of counterparty. We priced European options using the Black-Scholes model under a structural model, the first passage model which is a model for default risk. The formula for pricing European call and put options with default risks are given in equation (3) and equation (4).
Ignoring the default risk of the counterparty, we obtained the pricing formulae of European call and put options without default risk (Black-Scholes formula) in equation (5).
This pricing scheme can be used to derive the pricing formula for structured notes. As example, the derivation of the value of reverse exchangeable bond is given. The numerical result was given to confirm the validity of the pricing formula. In the numerical result, the tables and figures show the values of European call options with default risk and

Conclusion II

European options without default risk as $S(0), V(0), \sigma_v$ and T vary respectively. In figure 4.1,it was observed that as $S(0)$ increases, the value of European call option increases. As $V(0)$ increases, the value of a European option with default risk tends to the value of a European option without default risk. Also, as σ_v increases, the value of a European call option with default risk decreases. The values of European call option with default risk and European call option without default risk also increase as T varies.From the result,it is clear that the price of a European option with default risk is lower than the price of a European option without default risk.

The writer's assets value and underlying asset can also be modeled by a jump diffusion process.

Introducing a poisson process N with intensity λ under the probability measure $\mathbb{P}$ and a sequence of independent identically distributed random variables $(U_i)i \geq 1$ with finite expected value $\nu = \tilde{E}_{\mathbb{P}}(U_i)$.

Conclusion III

The firm's asset value and underlying asset are given by

$$\frac{dV(t)}{V(t)} = (r - \lambda\nu)dt + \sigma_v dW_v(t) + d\pi(t) \tag{8}$$

$$\frac{dS(t)}{S(t)} = (r - \lambda\nu)dt + \sigma_s dW_s(t) + d\pi(t) \tag{9}$$

where π is a jump process whose jump times are specified by the jump times of the poisson process N and the i^{th} jump is U_i.

$$\pi(t) = \sum_{i=1}^{N(t)} U_i \ \forall \ t \in [0, T]$$

References I

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*, 147–169.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481–2495.

Barré, P., Stöver, B. C., Müller, K. F., & Steinhage, V. (2017). Leafnet: A computer vision system for automatic plant species identification. *Ecological Informatics*, *40*, 50–56.

References II

Baweja, H. S., Parhar, T., Mirbod, O., & Nuske, S. (2018). Stalknet: A deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. In M. Hutter & R. Siegwart (Eds.), *Field and service robotics* (pp. 271–284). Cham: Springer International Publishing.

Behera, B., Kumaravelan, G., et al. (2019). Performance evaluation of deep learning algorithms in biomedical document classification. In *2019 11th international conference on advanced computing (icoac)* (pp. 220–224).

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Berlin, Heidelberg: Springer-Verlag.

References III

Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *Siam Review*, *60*(2), 223–311.

Boyd, S., Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Bryson, A. E., & Ho, Y. C. (1969). *Applied optimal control*. New York: Blaisdell.

Chellapilla, K., Puri, S., & Simard, P. (2006, October). High Performance Convolutional Neural Networks for Document Processing. In G. Lorette (Ed.), *Tenth International Workshop on Frontiers in Handwriting Recognition.* La Baule (France): Suvisoft. Retrieved from

References IV

https://hal.inria.fr/inria-00112631 (http://www.suvisoft.com)

Chen, J., Fan, Y., Wang, T., Zhang, C., Qiu, Z., & He, Y. (2018). Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. *Agronomy*, *8*(8), 129.

Chen, J., Liu, Q., & Gao, L. (2019). Visual tea leaf disease recognition using a convolutional neural network model. *Symmetry*, *11*(3), 343.

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 ieee conference on computer vision and pattern recognition* (pp. 3642–3649).

References V

Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010, December). Deep, big, simple neural nets for handwritten digit recognition. *Neural Comput.*, *22*(12), 3207–3220. Retrieved from `https://doi.org/10.1162/NECO_a_00052` doi: 10.1162/NECO_a_00052

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *CoRR*, *abs/1511.07289*.

Cooper, J., & Dobson, H. (2007, 09). The benefits of pesticides to mankind and the environment. *Crop Protection*, *26*, 1337-1348. doi: 10.1016/j.cropro.2007.03.022

References VI

Das, P., Acharjee, A., & Marium-E-Jannat. (2019). Double coated vgg16 architecture: An enhanced approach for genre classification of spectrographic representation of musical pieces. *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 1-5.

Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27, pp. 2933–2941). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2014/file/17e23e50bedc63b4095e3d8204ce063b-Paper.pdf

References VII

Dekel, O., Gilad-Bachrach, R., Shamir, O., & Xiao, L. (2012). Optimal distributed online prediction using mini-batches. *Journal of Machine Learning Research*, *13*(Jan), 165–202.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F. F. (2009, 06). Imagenet: a large-scale hierarchical image database. In (p. 248-255). doi: 10.1109/CVPR.2009.5206848

Denker, J. S., Gardner, W. R., Graf, H. P., Henderson, D., Howard, R. E., Hubbard, W., ... Guyon, I. (1989). Neural network recognizer for hand-written zip code digits. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 1* (pp. 323–331). Morgan-Kaufmann.

References VIII

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, *12*(7).

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New Yotk: John Willey & Sons.

Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2010, 06). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*, 303-338. doi: 10.1007/s11263-009-0275-4

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318.

References IX

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.

References X

Gaurav Belani. (2020). *Should you choose a GPU or a TPU to train your machine learning models.* https://www.predictiveanalyticsworld.com/machinelearningtimes/should-you-choose-a-gpu-or-a-tpu-to-train-your-machine-learning-models/10460/. ([Online; accessed 16-January-2020])

Girshick, R. B. (2015). Fast R-CNN. *CoRR*, *abs/1504.08083*. Retrieved from http://arxiv.org/abs/1504.08083

References XI

Glorot, X., Bordes, A., & Bengio, Y. (2011, 11–13 Apr). Deep sparse rectifier neural networks. In G. Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (Vol. 15, pp. 315–323). Fort Lauderdale, FL, USA: PMLR. Retrieved from `http://proceedings.mlr.press/v15/glorot11a.html`

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009, 06). A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, *31*, 855-68. doi: 10.1109/TPAMI.2008.137

References XII

Hahnloser, R., Sarpeshkar, R., Mahowald, M., Douglas, R., & Seung, H. (2000, 07). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*, 947-51. doi: 10.1038/35016072

He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the 2015 ieee international conference on computer vision (iccv)* (p. 1026–1034). USA: IEEE Computer Society. Retrieved from https://doi.org/10.1109/ICCV.2015.123 doi: 10.1109/ICCV.2015.123

References XIII

Hochreiter, S., & Schmidhuber, J. (1997, November). Long short-term memory. *Neural Comput.*, *9*(8), 1735–1780. Retrieved from https://doi.org/10.1162/neco.1997.9.8.1735 doi: 10.1162/neco.1997.9.8.1735

Howard, J., & Gugger, S. (2020, Feb). Fastai: A layered api for deep learning. *Information*, *11*(2), 108. Retrieved from http://dx.doi.org/10.3390/info11020108 doi: 10.3390/info11020108

Hu, J., Shen, L., & Sun, G. (2017). Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132-7141.

References XIV

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4700–4708).

Hubel, D. H., & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, *160*, 106–154.

Ioffe, S., & Szegedy, C. (2015, 07–09 Jul). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 448–456). Lille, France: PMLR.

References XV

Jähne, B., Scharr, H., Körkel, S., Jähne, B., Haußecker, H., & Geißler, P. (1999). Principles of filter design. In (Vol. 2, pp. 125–151). Academic Press.

Jarrett, K., Kavukcuoglu, K., Ranzato, M., & Lecun, Y. (2009, 09). What is the best multi-stage architecture for object recognition? In (Vol. 12). doi: 10.1109/ICCV.2009.5459469

Kaur, H., & Garg, H. (2014, 02). Pesticides: Environmental impacts and management strategies.. doi: 10.5772/57399

Kelley, H. J. (1960). Gradient theory of optimal flight paths. *ARS Journal*, *30*(10), 947-954.

References XVI

Khaki, S., Khalilzadeh, Z., & Wang, L. (2019). Classification of crop tolerance to heat and drought—a deep convolutional neural networks approach. *Agronomy*, *9*(12). Retrieved from https://www.mdpi.com/2073-4395/9/12/833 doi: 10.3390/agronomy9120833

Kim, K.-H., Kabir, E., & Jahan, S. A. (2017). Exposure to pesticides and the associated human health effects. *Science of The Total Environment*, *575*, 525 - 535. Retrieved from http://www.sciencedirect.com/science/article/pii/S004896971631926X doi: https://doi.org/10.1016/j.scitotenv.2016.09.009

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

References XVII

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. *CoRR*, *abs/1706.02515*. Retrieved from `http://arxiv.org/abs/1706.02515`

Knillmann, S., & Liess, M. (2019, 01). Pesticide effects on stream ecosystems: Drivers, risks, and societal responses. In (p. 211-214). doi: 10.1007/978-3-319-96229-0_33

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. *arXiv preprint arXiv:1404.5997*.

Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012, 01). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, *25*. doi: 10.1145/3065386

References XVIII

LeCun, Y. (1985). Une procedure d'apprentissage pour reseau a seuil asymmetrique (a learning scheme for asymmetric threshold networks). In *Proceedings of cognitiva 85, paris, france* (pp. 599–604).

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Back-propagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541-551.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the ieee* (Vol. 86, pp. 2278–2324). Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.7665`

References XIX

LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the 2004 ieee computer society conference on computer vision and pattern recognition* (p. 97–104). USA: IEEE Computer Society.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the ieee international conference on computer vision* (pp. 2980–2988).

Liu, C., Yin, F., Wang, Q., & Wang, D. (2011). ICDAR 2011 Chinese handwriting recognition competition. In *11th international conference on document analysis and recognition* (p. 1464-1469).

References XX

Long, J., Shelhamer, E., & Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *CoRR*, *abs/1411.4038*. Retrieved from `http://arxiv.org/abs/1411.4038`

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *International conference on machine learning (icml)*. Retrieved from `http://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf`

McCulloch, W. S., & Pitts, W. (1943, Dec 01). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

References XXI

Mohanty, S., Hughes, D., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. *arXiv preprint arXiv:1604.03169*.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on international conference on machine learning* (p. 807–814). Madison, WI, USA: Omnipress.

Nemirovski, A., Juditsky, A., Lan, G., & Shapiro, A. (2009). Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, *19*(4), 1574–1609.

References XXII

Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence o(1/k2). In *Doklady ansssr* (Vol. 269, pp. 543–547).

Oh, K.-S., & Jung, K. (2004). GPU implementation of neural networks. *Pattern Recognition*, *37*(6), 1311–1314.

Oliveira, I., Cunha, R. L., Silva, B., & Netto, M. A. (2018). A scalable machine learning system for pre-season agriculture yield forecast. *arXiv preprint arXiv:1806.09244*.

Parr, T., & Howard, J. (2018). The matrix calculus you need for deep learning. *CoRR*, *abs/1802.01528*. Retrieved from `http://arxiv.org/abs/1802.01528`

References XXIII

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *The Official Journal of the International Neural Network Society*, *12*(1), 145–151.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Swish: a self-gated activation function. *CoRR*, *abs/1710.05941*. Retrieved from `http://arxiv.org/abs/1710.05941`

References XXIV

Rasmussen, C. B., & Moeslund, T. B. (2019). Maize silage kernel fragment estimation using deep learning-based object recognition in non-separated kernel/stover rgb images. *Sensors*, *19*(16), 3506.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400–407.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65–386.

References XXV

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & P. R. Group (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 1: foundations, pp. 318–362). Cambridge, MA, USA: MIT Press.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Li, F. F. (2014, 09). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*. doi: 10.1007/s11263-015-0816-y

References XXVI

Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., ... Siegwart, R. (2018). Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, *10*(9), 1423.

Sanborn, M., Kerr, K., Sanin, L., Cole, D., Bassil, K., & Vakil, C. (2007, 10). Non-cancer health effects of pesticides systematic review and implications for family doctors. *Canadian family physician Médecin de famille canadien*, *53*, 1712-20.

Sanchez-Bayo, F., & Goka, K. (2014, 04). Pesticide residues and bees - a risk assessment. *PloS one*, *9*, e94482. doi: 10.1371/journal.pone.0094482

References XXVII

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... Hassabis, D. (2016, January). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. doi: 10.1038/nature16961

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition.*

References XXVIII

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, *2016*.

Sobel, I., & Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing. *Stanford Artificial Intelligence Project*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929-1958. Retrieved from `http://jmlr.org/papers/v15/srivastava14a.html`

References XXIX

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2011, 09). The german traffic sign recognition benchmark: A multi-class classification competition. In (p. 1453 - 1460). doi: 10.1109/IJCNN.2011.6033395

Stephen, O., Sain, M., Maduh, U. J., & Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, *2019*.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (1st ed.). Cambridge, MA, USA: MIT Press.

References XXX

Syngenta. (2019). *Syngenta Crop Challenge In Analytics.* `https://www.ideaconnection.com/syngenta-crop-challenge/challenge.php`. ([Online; accessed 16-January-2020])

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1–9).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2818–2826).

References XXXI

Taghavi, S., Esmaeilzadeh, M., Najafi, M., Brown, T., & Borevitz, J. (2018, 12). Deep phenotyping: Deep learning for temporal phenotype/genotype classification. *Plant Methods*, *14*. doi: 10.1186/s13007-018-0333-4

Tao, A., Barker, J., & Sarathy, S. (2016). Detectnet: Deep neural network for object detection in digits. *Parallel Forall*, *4*.

Tetila, E. C., Machado, B. B., Menezes, G. V., de Souza Belete, N. A., Astolfi, G., & Pistori, H. (2019). A deep-learning approach for automatic counting of soybean insect pests. *IEEE Geoscience and Remote Sensing Letters*.

References XXXII

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1990). Phoneme recognition using time-delay neural networks. In *Readings in speech recognition* (p. 393–404). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences* (Unpublished doctoral dissertation). Harvard University.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, *1*.

References XXXIII

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1492–1500).

Yu, J., Sharpe, S. M., Schumann, A. W., & Boyd, N. S. (2019). Deep learning for image-based weed detection in turfgrass. *European journal of agronomy*, *104*, 78–84.

Zagoruyko, S., & Komodakis, N. (2017). *Wide residual networks.*

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). *Dive into deep learning*. (`https://d2l.ai`)

Zhang, Z. (2019). Deep learning for field-based automated high-throughput plant phenotyping. *Graduate Theses and Dissertations*, *17628*.

THANKS

THANK YOU FOR LISTENING