Detecting the Invasive Japanese Honeysuckle With Machine Learning

Word Count: 4519

Abstract

With the rise of globalization, there has been an increased spread of invasive plant species to the United States. One particularly damaging invasive species is named the Japanese Honeysuckle (Lonicera japonica), which is a perennial woody vine with white flowers and black fruit. The East Asian vine was introduced to the United States in the 1800s as a landscape plant but quickly spread across the east and gulf coasts due to its biologically advantageous traits. Some of these traits include being aggressive and the ability to reproduce through several methods. In this study, a traditional convolutional neural network (CNN) and VGG16 CNN architecture were used to detect the vine in a dataset of test images. Moreover, these machine learning models were trained with images that contained the vine displaying its flowering traits and images that included the vine without said traits. In addition to that, some of the images that were used for training did not include the vine at all. Overall, the VGG16 model trained on photos with the traits present had the best accuracy with 94 percent, while the traditional CNN trained on the images without the traits had the lowest accuracy with 63 percent. However, all of the algorithms had an accuracy of more than 50 percent, which suggested that they were all detecting the vine to some extent in the test dataset. The results of this study supported that it was, in fact, feasible to detect the Japanese Honeysuckle with CNN machine learning methods to a large extent in a dataset of images collected from the gulf coast. In the end, this study not only confirms other research in this field but could also help conservation agencies build automated systems to detect the invasive Japanese Honeysuckle in the future.

Keywords: Japanese Honeysuckle, Invasive Species, Machine Learning, CNN, VGG16

Detecting the Invasive Japanese Honeysuckle With Machine Learning

Introduction

As the world becomes ever more interconnected, there is a greater risk of invasive species being unintentionally introduced to areas that are unprepared and ecologically vulnerable to their attacks (Early et al., 2016). According to the U.S. Department of Agriculture, invasive species are "plants, animals, and other living organisms" that cause harm and are not native to the land or body of water that they invade (United States Department of Agriculture, 2019). These organisms cost the North American economy tens of billions of dollars every year. Moreover, recent data even suggest that this staggering cost is expected to grow in the future. Some of the causes of this enormous drain on the economy are the damage these species do to the North American agriculture, infrastructure, and real estate sectors (Crystal-Ornelas et al., 2021). However, it should be noted that only invasive plant species will be the primary research focus of this study.

Like most other invasive species, invasive plants cause not only fiscal harm but also ecological damage to the environments they invade by outcompeting native species and changing said area's biodiversity (Kerns, 2012; Crystal-Ornelas et al., 2021). Furthermore, studies in this field also support the idea that highly invasive species are able to directly outcompete native species by either having more biologically advantageous traits such as the ability to resist local consumers or capture more of the environment's available resources. These traits ultimately allow said plant species to collect more nutrients, water, sunlight, or space than their native counterparts. This increased competition can subsequently lead to the decline of the native species in a given area (Gioria & Osborne, 2014; Profetto, 2021). In addition to that, invasive species can also indirectly compete with native species by causing consumers to target them more for predation. This is best shown by the relationship between the invasive grass species named the European Marram Grass (Ammophila Arenaria) and an endangered Californian plant called the Lupine (Lupinus tidestromii). In this relationship, the invasive grass attracts more small rodents, which eat the seeds of the native grass. In the end, this phenomenon has helped contribute to the decline of the Lupine (Dangremond et al., 2010). Moreover, the effect of some invasive plant species' ability to outcompete native species either directly or indirectly can also result in the decline of biodiversity in a given area. Data suggests that this is primarily due to the fact that invasive species can often displace native ones (Gioria & Osborne, 2014). With all this information in mind, some reports suggest that invasive species have contributed to the extinction of several species in the past (Gurevitch et al., 2004).

While preventing invasive plant species from being introduced to a given area is the most effective form of control, current research suggests that there are several other ways to combat invasive species on a smaller scale and larger scale. These methods include simply physically removing them from the area and using herbicides (New England Wild Flower Society, 2006). Regarding the use of chemical forms of mass plant control such as herbicides, the academic consensus seems to be that their use is effective but comes with some very important caveats. Some disadvantages that using chemical herbicides presents are that it can damage the environment, native species, and the health of humans exposed to it (Nicolopoulou-Stamati et al., 2016). In addition to that, simply identifying an area that contains invasive species that need to be removed via the processes above requires conservationists or scientists to survey the area of land, which can cost both time and money (National Park Service, 2018). Ultimately, this problem of quickly monitoring and subsequently removing invasive plant species on a broader scale has contributed to the use of machine learning programs to survey areas of invaded land. As explained by IBM, an American technology company, machine learning is a subset of artificial intelligence, which focuses on providing programs with large amounts of data so that they may improve upon their accuracy when presented with similar data sets or scenarios (IBM Cloud Education, 2020). Once trained on said data sets collected by crewless aerial vehicles, people, or other methods, these algorithms, in some cases, can detect the target items or more relevantly invasive plants with a high degree of accuracy. Ultimately, these initial findings narrowed this paper's research focus to the field of detecting invasive plant species with machine learning.

Literature Review

With the initial research above in mind, several sources were reviewed in this field to create and sharpen this paper's research focus into a research question. One particularly important source was (Jensen et al., 2020). In this study, the researchers detected the invasive Kudzu vine with several machine learning methods with different data types in order to see which combination produced the best results. For context, the Kudzu vine (Pueraria Montana) is a semi-woody vine that is native to East Asia. It first arrived in the United States in 1876 and then again in subsequent years as an ornamental plant due to its aesthetic appeal. Some of these alluring traits that enticed past Americans were the vine's wide leaves and purple flowers that appear in the late summertime. While the initial importation of this vine did allow it to spread throughout limited locations in the country, the use of this vine as a soil erosion protectant, livestock feed, and soil-enriching agent during the Great Depression sped up its invasion exponentially (Profetto, 2021). Once established in an area, this vine's speedy growth and ability to climb over native species contributed to its dominance in the south. In the research paper, the researchers collected data in Atlanta, Georgia, on the vine in several electromagnetic spectrum ranges to train machine learning methods. It is also explained that data was gathered from several sources like Google Earth and satellites. After training and testing their methods, they concluded that the Random Forest, Neural Network, and Support Vector Machine algorithms were the best performing overall. It was also noted that using a combination of data was the best way to get the most effective results. With regards to this study, the research paper explained above not only confirmed the assumption that detecting invasive vines were possible but advocated that there was a need for more research in this subfield. This gap in the aforementioned subfield was supported by several other papers in this field and helped develop the criterion used to select the research subject outlined later in this paper (Jensen et al., 2020; Aneece & Epstein, 2015; Gaston, 2018).

Another influential study that was pivotal to the creation of this investigation was named (Ashqar & Abu-Naser, 2019) by researchers at Al-Azhar University-Gaza, Palestine. In this relatively short but impactful paper, the researchers used 2295 images of an invasive hydrangea species in different scenarios to train a machine learning method named a convolutional neural network or CNN. In an article by Thomas Wood, it is explained that "A convolutional neural network, or CNN, is a deep learning neural network designed for processing structured arrays of data such as images" (Thomas Wood, 2019). Moreover, this type of machine learning method finds patterns in images in order to identify them. The ingeniousness of this method is revealed by its construction. In summation, a CNN is composed of several layers that include convolutional and pooling layers. The convolutional layer searches a given image for patterns while the pooling layers break an image into smaller subsections that the convolutional layer can then use to look for more nuanced features. In the paper, the researchers used a special type of pretrained CNN architecture named VGG16 since its design made it highly effective at the task that their research was trying to complete. With this in mind, they were able to get an accuracy of over 99.7 percent on their best model. Overall, the results of this paper and that of others similar to it further supported that detecting invasive plant species was possible and that using a CNN and VGG16 architecture for other studies in this field would be effective (Ashqar & Abu-Naser, 2019; Rodriguez et al., 2021; Dhaka et al., 2021).

The last major study that will be reviewed in this literary review is named (Alvarez-Taboada et al., 2017) by researchers in Europe. In this study, the researchers investigated detecting the invasive Hakea Sericea tree species that has invaded Europe and, more relevantly to their research Portugal. The Hakea Sericea tree is native to south-eastern Australia and has led to a reduction in the abundance of native plant species and increased chances of forest fires in invaded areas. In order to detect the vine, the researchers implemented a highly nuanced and technical method that compared the results of two nearest neighbor machine learning programs with both worldview satellite data and unmanned aerial vehicle (UAV) data. Their results were shown in the form of maps displaying the general distribution of the vine in a given area. It should be noted that they used UAV data, which was somewhat uncommon among the reviewed sources. However, they justified its use by explaining that the vehicles used to collect the data were reflective of what would be used for automated detection systems for the invasive Hakea Sericea in the future. It was also explained that the use of these UAVs could decrease the time and cost of surveying an area of land for this species. In the end, they were able to achieve an accuracy in the 90 percent range for their satellite data-powered method, while the UAV data got results in the 70 percent range. While their results and method were very helpful in guiding this paper's method, the researchers' sub-conclusion was the most impactful to this paper. This primarily was because they explained that using the distinctive traits of an invasive plant during its flowering period could be used to detect it better. With this taken into consideration, this paper's method section was improved to better answer the research question explained in the next section.

Research Subject

Although research in this field has addressed the detection of several other highly invasive vines and plants, after reviewing sources, there seemed to be a lack of research detecting the Japanese honeysuckle vine, which was selected as the subject of this research paper. This selection was based on the fact that this vine had very little research about it in this field, was invasive, and displayed distinctive traits that would make it possible to detect with machine learning. The Japanese Honeysuckle (Lonicera japonica) is a perennial woody vine that produces white flowers and black fruit. Moreover, this vine is also considered to be a part of the "dirty dozen" most invasive plant species by many state-run conservation agencies (Kurtz et al., 2015; Coffey, 2021). The leaves of this vine are oval-shaped and are usually paired on each side of the vine. This vine was first introduced to the United States in 1806 on Long Island, NY, from East Asia (Kurtz et al., 2015). When introduced to the United States, studies suggest that the vine spread quickly because of its aggressive nature and various methods of reproduction. These methods include producing seeds, independent underground rhizomes, and runners created from separated sections of the mother vine (Kurtz et al., 2015). As of the creation of this paper, this vine primarily inhabits the east and southeast coast; however, studies support that its range could expand due to global warming, which would increase the economic and ecological damage done by this invasive species (Wang et al., 2012). Currently, methods of combating this vine's growth include directly surveying land to remove the vine or to use herbicides. However, research suggests that these methods are only effective for short periods and need to be used frequently, which can cause several negative consequences (New England Wild Flower Society, 2006; Kurtz et al., 2015). In the end, this confluence of factors led to the creation of this paper's focus, answering the research question of to what extent using machine learning is feasible in detecting the Japanese Honeysuckle from images.

Method

In this paper, an experimental design method was used in order to adequately address the research question and stay consistent with numerous other papers in this field. The use of this method is necessary to answer this paper's research question since there is very little data on detecting the Japanese Honeysuckle with machine learning. This means that a review of the accuracies of other researchers detecting this vine can not be used to perform a content analysis, as of the creation of this paper. This same issue of there being very few studies detecting one plant species could be a contributing factor as to why most related sources in this field also implement an experimental design method. For this study, images of the Japanese Honeysuckle will be collected to be used in a traditional CNN and VGG16 machine learning model to produce accuracy values that will later be compared. Furthermore, the images of the vine will be broken into three main groups: flowering traits of the vine shown, flowering traits of the vine not shown, and vine not present.

Assumptions and Hypotheses

After considering the results of the aforementioned papers in the literary review and that of others in this field, a hypothesis was formed that was consistent with the results of others in this field. The hypothesis was that it is, in fact, feasible to a large extent to detect the invasive

Japanese Honeysuckle vine with CNN and VGG16 machine learning models with the test dataset. Moreover, it was also initially hypothesized that VGG16 would produce better results than the traditional CNN method for all datasets. Specifically, it was hypothesized that the VGG16 model with the phenological or flowering traits of the vine present in the data would perform the best overall. It should be noted that the second more detailed hypothesis of the accuracy of the VGG16 model was generated with several assumptions about this research process. Some of these assumptions included predicting that it would be possible to collect enough data to build a testable model and that the VGG16's pre-trained parameters would be effective in detecting the Japanese Honeysuckle. While blindly assuming these things could pose issues later in the research process, the studies in the literature review and broader field in general quickly proved these assumptions to be accurate and justified.

Data Collection

For this method, 1500 total images were collected in order to feed into the CNN and VGG16 machine learning methods. These images were collected from the global biodiversity information facility (GBIF) and iNaturalist, which are both platforms that have vast datasets on thousands of species from around the world. In specific, iNaturalist is a crowdsourced platform that allows anyone to take images of flora and fauna to upload to their site. After a given amount of time, some of these images are confirmed by the scientific community and then subsequently labeled as research-grade. From there, the image and its accompanying metadata are input into GBIF's collective database. Moreover, the label research-grade means that the uploaded image was identified as the correct species. Only research-grade images were used to build, train, and test the machine learning models since their unique distinction ensured that they were correctly labeled. These given images were also only of plant occurrences in the gulf coastal region of the

United States in the year 2021 to reflect the current range of the invasive vine in addition to limiting the number of searchable images to comb through. While this range is comparatively large, it is necessary since there were very few images of the Japanese Honeysuckle compared to other species in the searchable database of images from GBIF. This means that a bigger area was needed to gather enough images to build and train the models. Concerning their layout, the images of the vine were broken into three main groups: flowering traits of the vine shown, flowering traits of the vine not shown, and vine not present. Each group had 500 images. For the images where the vine was not present, the GBIF dataset was also used with the same search parameters to gather them. These images were composed of species that could be located near the Japanese Honeysuckle. This was done in order to ensure that the results produced from the dataset were reflective of what an automated detection system would encounter in the real world. These three general groups of images were then split into the train, valid, and test data sets and paired so that they could be inputted into the machine learning methods. These sets include 300, 100, and 100 images, respectively.

Preprocessing

Google Colab was used to construct the CNN and VGG16 models. Google Colab is a platform built by Google to allow individuals to build applications and projects in Python, which is a common computer science language used for machine learning purposes. This application was used because it is not resource-intensive, connects to one's google drive, and allows the entire project to be easily accessed through a link. This last feature is especially important since it allows other researchers to easily access this paper's code and build upon the results of this study. In addition to that, it should be noted that Keras and the TensorFlow models were used for this project since they provide many of the necessary functions to build machine learning methods. Both the traditional CNN and VGG16 methods first preprocess the images so that the said algorithms could use them. This process first started with the creation of folders in the connected google drive that were assigned the names outlined in the data collection section. Within these folders, subfolders named "Honeysuckle" and "Nothoneysuckle" were created to hold images. Figure (1) shows how the folders and images were distributed. After that, both programs converted the images into NumPy arrays with a 244 by 244-pixel size and a binary class setup. This pixel size was selected since it is necessary for the VGG16 model. Once this action was performed by both methods, they checked to see whether this process was done correctly and then printed some of the newly formatted images to the console. One example of the output is shown in figure (2). The color distortion seen in the figure is due to the conversion of the RGB color data to a float during printing and does affect the results of the machine learning algorithms.

CNN Model

After the images were preprocessed, the CNN machine learning method initialized its structure. This process involved the creation of a sequential model. This type of model was used since, as explained by TensorFlow on their website, it allowed for the creation of stackable layers that output a given tensor or value (TensorFlow, 2022). The model used filter values between 16 and 64 that doubled in size for the convolutional layers. These filters are what give the convolutional layers the ability to find patterns in the images (Thomas Wood, 2019). These sizes were used since research supports starting off with a smaller filter size and then increasing it with each layer yields a better result for CNN machine learning applications (Camgözlü & Kutlu, 2020). Furthermore, it should also be noted that the Kernel size was set to three by three and that the activation function Relu was selected for the layer parameters. These are relatively standard parameters that can be found on most other models using Keras; however, it should be noted that the final layer had an output of one and an activation function of sigmoid since this CNN used binary cross-entropy. In summation, this means that the program produces the output of 0 for Honeysuckle and 1 for Nothoneysuckle. After the model was built, the model.summary() function revealed that the method had trainable 81,185 parameters to learn how to identify how to detect the Japanese Honeysuckle during the training and validation phases. During these phases, the model went through all of the images in the dataset 25 times or epochs to train and validate itself. Due to time and computing constraints, this relatively small number of epochs was chosen. Once the model was trained, the model.predict() function was used to allow the algorithm to make predictions about the images in the test dataset. Due to the set up of the method, the CNN produced a binary output as explained above. The results were then printed and compared, which is further explained in the results section.

VGG16 Model

Due to the fact that the VGG16 model was a pretrained model created and optimized by the Visual Geometry Group Lab of Oxford University for a challenge, its implementation was more straightforward compared to the traditional CNN architecture (pawangfg, 2020). After preprocessing, the VGG16 model simply had to be imported directly into the Google Colab notebook running the project. From there, the model.summary() function revealed that the method had a staggering 138,357,544 trainable parameters and 0 non-trainable ones. Unfortunately, this was simply too many parameters to use for this application, so 4,097 of them were converted into trainable parameters, and 134,260,544 were changed to non-trainable parameters. After that, the model was trained on both sets of train and valid images for five epochs. Like the CNN model, the number of epochs was limited in order to train the method in a reasonable amount of time, given the constraints of using Google Colab. After the model was trained, the model.predict() function was used to get the binary predictions of the method for the test images. These predictions were then printed and compared in the results and conclusion section.

Results & Conclusions

After the CNN and VGG16 models were built, trained, validated, and tested, the results were printed in two ways. The first way was through a confusion matrix which is the standard for understanding the results of a given machine learning method. As shown in figure (3), there is the predicted label axis at the bottom of the confusion matrix, and on the left side, there is the true label axis. When both of these axes meet in the top left and bottom right squares are instances where the algorithm correctly predicted an image as the invasive vine or not. The other squares not stated are the instances where the machine learning method was incorrect in its prediction. Lastly, in order to calculate the programs' accuracy, the total number of correct instances is divided by the total size of the test database. The second way that the results were printed was through graphing a ROC curve which displays each algorithm's false positive and true positive rate on each axis. Overall, it can be summarized that the more area under the curve for this graph, the better the method is at detecting the intended output. It should be noted that these output programs were built with the help of free resources from Scikit-learn. After all of the data was tabulated, it was shown that the VGG16 model with the images of the flowering vine had the best results with an accuracy of about 94 percent, while the traditional CNN model with images of the vine not flowering had the lowest accuracy of about 63 percent. All of the results are shown in Table (1). Overall, the prediction that the VGG16 models would outperform their counterparts was correct and shown by the large difference in their accuracy between the traditional CNN models. This could be due to the sheer size of the VGG16 model and the fact that it contained pretrained parameters that could have already been able to detect some aspects of the vine in the datasets. Furthermore, it was also supported that the machine learning method that used images of the flowering vine performed better than those that did not. This ultimately supported the conclusion explained in (Alvarez-Taboada et al., 2017) that using an invasive species' phenological traits could improve its chances of being detected, and thus detecting invasive plants during the spring would be easier with current tactics. Overall, the results from all of the machine learning methods across all of the datasets showed that the methods were, in fact, able to detect the vine better if they were guessing. This means that the results of the experiment support the conclusion that it is feasible to a large extent to detect the Japanese Honeysuckle vine with traditional and pretrained CNN-based approaches in the gulf coast region of the United States.

Limitations

Due to the nature of machine learning applications, there were several limitations to the method used and the results that were produced from said method. The first and biggest limitation that should be addressed is the structure of the models. While the models built were consistent with others in the field, they could have been better fine-tuned to account for the overfitting shown in some of the ROC curves in the appendix. Overfitting is when there is a sizable drop in accuracy from training to testing because the machine learning algorithm has not generalized enough (IBM Cloud Education, 2021). With more resources, future researchers could account for this overfitting by optimizing the structure and finding the optimal amount of epochs to train each method for. Another limitation of this study is the data size used to train the method since most machine learning applications use tens of thousands of images. Unfortunately, this could not be done with this vine in this instance due to the limited amount of free images of it in the gulf coastal area. With access to a larger amount of data, studies suggest that the final accuracy of the machine learning methods would be similar, if not higher (Althnian et al., 2021). While these limitations should be considered when reflecting on the results of this study, they do not invalidate the conclusions provided since research into this field supports that if these limitations were to be addressed, the results would only be more concrete instead of wildly varying.

Discussion

Regarding this study's relevance, it first serves as a confirmation of the (Alvarez-Taboada et al., 2017) conclusion, which suggested using the flowering traits of a species to better detect it. In addition to supporting the findings of other researchers in this field, this study could also serve as a starting point for other researchers looking to detect this vine in other areas of the United States with other machine learning methods not used in this study. This is primarily because there is still a need for more research to detect this vine in more complex situations. Last but not least, this study could help conservation agencies and state governments to develop autonomous detection systems like those explained in (Alvarez-Taboada et al., 2017; Rodriguez et al., 2021) but for the gulf coastal region. Once implemented, this system could allow for more direct targeting of the Invasive Japanese Honeysuckle to curb its spread. As explained in the introduction, these systems would also cut down on the cost and amount of human resources needed to fight these vines by allowing agencies to work smarter, not harder.

References

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. https://doi.org/10.1109/icengtechnol.2017.8308186

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), 796. https://doi.org/10.3390/app11020796

Alvarez-Taboada, F., Paredes, C., & Julián-Pelaz, J. (2017). Mapping of the Invasive Species Hakea sericea Using Unmanned Aerial Vehicle (UAV) and WorldView-2 Imagery and an Object-Oriented Approach. *Remote Sensing*, *9*(9), 913. https://doi.org/10.3390/rs9090913

Aneece, I., & Epstein, H. (2015). Distinguishing Early Successional Plant Communities Using Ground-Level Hyperspectral Data. *Remote Sensing*, *7*(12), 16588–16606. https://doi.org/10.3390/rs71215850

Ashqar, B. A. M., & Abu-Naser, S. S. (2019, March 1). *Identifying Images of Invasive Hydrangea Using Pre-Trained Deep Convolutional Neural Networks*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3369016

Camgözlü, Y., & Kutlu, Y. (2020). *Analysis of Filter Size Effect In Deep Learning*. https://arxiv.org/pdf/2101.01115.pdf

Coffey, S. (2021, April 12). *Japanese Honeysuckle: A "Dirty Dozen" Plant Lewis Ginter Botanical*. Lewis Ginter Botanical Garden. https://www.lewisginter.org/japanese-honeysuckle/

Crisigiovanni, E. L., Filho, A. F., Pesck, V. A., & Lima, V. A. de. (2020). *The Effective Use of Machine Learning in WorldView-2 Image for the Recognition of Two Forest Species in the Brazilian Atlantic Rainforest*. https://doi.org/10.21203/rs.3.rs-33722/v1

Crystal-Ornelas, R., Hudgins, E. J., Cuthbert, R. N., Haubrock, P. J., Fantle-Lepczyk, J., Angulo, E., Kramer, A. M., Ballesteros-Mejia, L., Leroy, B., Leung, B., López-López, E., Diagne, C., & Courchamp, F. (2021). Economic costs of biological invasions within North America. *NeoBiota*, *67*(10.3897), 485–510. https://doi.org/10.3897/neobiota.67.58038

Dangremond, E. M., Pardini, E. A., & Knight, T. M. (2010). Apparent competition with an invasive plant hastens the extinction of an endangered lupine. *Ecology*, *91*(8), 2261–2271. https://doi.org/10.1890/09-0418.1

Demertzis, K., & Iliadis, L. (2016). Detecting invasive species with a bio-inspired semi-supervised neurocomputing approach: the case of Lagocephalus sceleratus. *Neural Computing and Applications*, *28*(6), 1225–1234. https://doi.org/10.1007/s00521-016-2591-2

Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Kavita, K., Ijaz, M. F., & Woźniak, M. (2021). A Survey of Deep Convolutional Neural Networks Applied for Prediction of Plant Leaf Diseases. *Sensors*, *21*(14), 4749. https://doi.org/10.3390/s21144749

Early, R., Bradley, B. A., Dukes, J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., Gonzalez, P., Grosholz, E. D., Ibañez, I., Miller, L. P., Sorte, C. J. B., & Tatem, A. J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms12485 freeCodeCamp.org. (2020, June 18). *Keras with TensorFlow Course - Python Deep Learning and Neural Networks for Beginners Tutorial*. Www.youtube.com. https://www.youtube.com/watch?v=qFJeN9V1ZsI

Gaston, K. J. (2018). UAVs and Machine Learning Revolutionising Invasive Grass and Vegetation Surveys in Remote Arid Lands. *Sensors*, *18*(2), 605. https://doi.org/10.3390/s18020605

GBIF.org (10 April 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.58gfs4

GBIF.org (10 April 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.a9knqm

GBIF.org (29 April 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.phfa97

Gioria, M., & Osborne, B. A. (2014). Resource competition in plant invasions: emerging patterns and research needs. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00501

Gurevitch, J., Padilla, D., & Blackburn, T. (2004). Mammal extinctions on Australian islands: causes and conservation implications. *Committee on Recently Extinct Organisms)*, *19*, 5446–5451. https://doi.org/10.1016/j.tree.2005.01.003

IBM Cloud Education. (2020, July 15). *What is Machine Learning?* Www.ibm.com; IBM. https://www.ibm.com/cloud/learn/machine-learning

IBM Cloud Education. (2021, March 3). *What is Overfitting?* Www.ibm.com. https://www.ibm.com/cloud/learn/overfitting

Jensen, T., Seerup Hass, F., Seam Akbar, M., Holm Petersen, P., & Jokar Arsanjani, J. (2020). Employing Machine Learning for Detection of Invasive Species using Sentinel-2 and AVIRIS Data: The Case of Kudzu in the United States. *Sustainability*, *12*(9), 3544. https://doi.org/10.3390/su12093544

Keras. (2020). *The Sequential model*. Keras.io. https://keras.io/guides/sequential_model/

Kerns, B. (2012, September). *Invasive Plants in Forests and Rangelands | Climate Change Resource Center*. Www.fs.usda.gov; U.S. Department of Agriculture, Forest Service, Climate Change Resource Center. https://www.fs.usda.gov/ccrc/topics/invasive-plants

Kurtz, Cassandra M.; Hansen Mark H. 2015. An assessment of Japanese honeysuckle in northern U.S. forests. Res. Note NRS-202. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 5 p

National Park Service. (2018, December 6). *Invasive Plant Species Monitoring (U.S. National Park Service)*. Www.nps.gov. https://www.nps.gov/im/medn/invasive-plants.htm

New England Wild Flower Society. (2006). *Managing Invasive Plants Methods of Control* (pp. 21–23). New England Wild Flower Society.

Nicolopoulou-Stamati, P., Maipas, S., Kotampasi, C., Stamatis, P., & Hens, L. (2016). Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture. *Frontiers in Public Health*, *4*. https://doi.org/10.3389/fpubh.2016.00148

pawangfg. (2020, February 26). *VGG-16 | CNN model*. GeeksforGeeks. https://www.geeksforgeeks.org/vgg-16-cnn-model/

Profetto, G. (2021). *Kudzu invasion and control in southern upland forests of Kudzu invasion and control in southern upland forests of Mississippi Mississippi*. https://scholarworks.uno.edu/cgi/viewcontent.cgi?article=4174&context=td

Rodriguez, R., Perroy, R. L., Leary, J., Jenkins, D., Panoff, M., Mandel, T., & Perez, P. (2021). Comparing Interpretation of High-Resolution Aerial Imagery by Humans and Artificial Intelligence to Detect an Invasive Tree Species. *Remote Sensing*, *13*(17), 3503. https://doi.org/10.3390/rs13173503

Sabottke, C. F., & Spieler, B. M. (2020). The Effect of Image Resolution on Deep Learning in Radiography. *Radiology: Artificial Intelligence*, *2*(1), e190015. https://doi.org/10.1148/ryai.2019190015

Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

Szilassi, P., Szatmári, G., Pásztor, L., Árvai, M., Szatmári, J., Szitár, K., & Papp, L. (2019). Understanding the Environmental Background of an Invasive Plant Species (Asclepias syriaca) for the Future: An Application of LUCAS Field Photographs and Machine Learning Algorithm Methods. *Plants*, *8*(12), 593. https://doi.org/10.3390/plants8120593

TensorFlow. (2022). *The Sequential model | TensorFlow Core*. TensorFlow. https://www.tensorflow.org/guide/keras/sequential_model

Thambawita, V., Strümke, I., Hicks, S. A., Halvorsen, P., Parasa, S., & Riegler, M. A. (2021). Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images. *Diagnostics*, *11*(12), 2183. https://doi.org/10.3390/diagnostics11122183

The City of Portland Oregon. (2010, December 17). *The Problem With Invasive Plants | About Invasive Plants | The City of Portland, Oregon*. Portlandoregon.gov. https://www.portlandoregon.gov/bes/article/330681

Thomas Wood. (2019, May 17). *Convolutional Neural Networks*. DeepAI. https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network

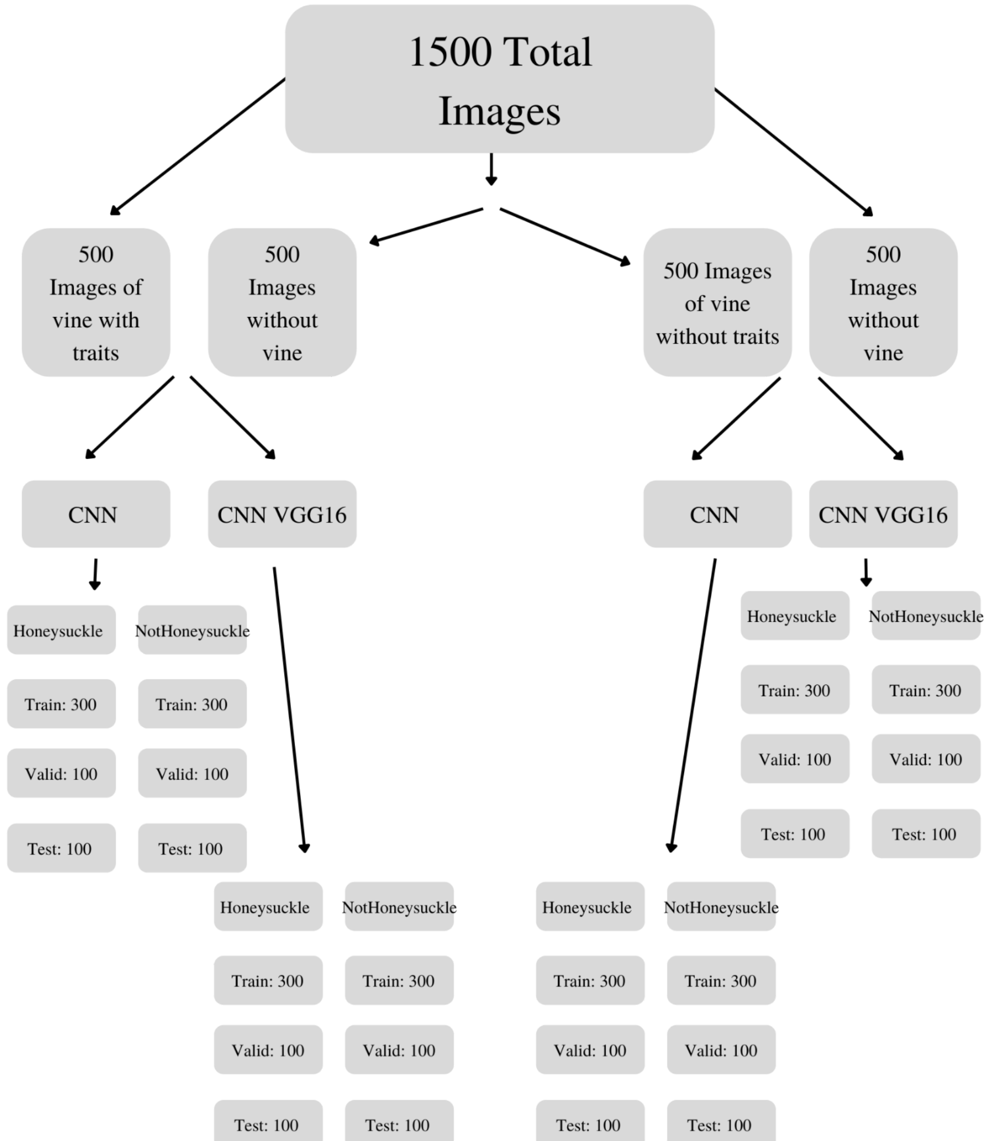
Verma, M. (2020, November 13). *Binary Classification Using Convolution Neural Network (CNN) Model*. Medium. https://medium.com/@mayankverma05032001/binary-classification-using-convolution-neural-network-cnn-model-2635ddcdc510

Wang, H.-H., Wonkka, C. L., Grant, W. E., & Rogers, W. E. (2012). Potential Range Expansion of Japanese Honeysuckle (Lonicera japonica Thunb.) in Southern U.S. Forestlands. *Forests*, *3*(3), 573–590. https://doi.org/10.3390/f3030573

Wang, Z., Schaaf, C. B., Sun, Q., Kim, J., Erb, A. M., Gao, F., Román, M. O., Yang, Y., Petroy, S., Taylor, J. R., Masek, J. G., Morisette, J. T., Zhang, X., & Papuga, S. A. (2017). Monitoring land surface albedo and vegetation dynamics using high spatial and temporal resolution synthetic time series from Landsat and the MODIS BRDF/NBAR/albedo product. *International Journal of Applied Earth Observation and Geoinformation*, *59*, 104–117. https://doi.org/10.1016/j.jag.2017.03.008

Figure 1

Image distribution flow chart



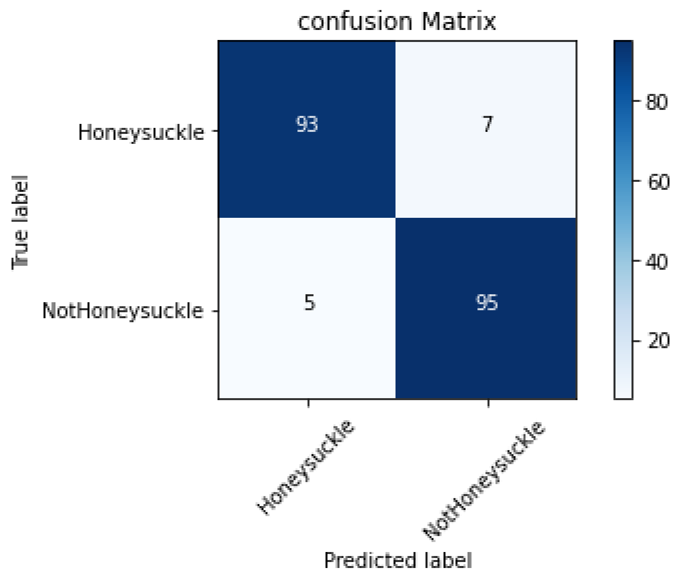
Note: This flow chart provides the distribution of images throughout the creation of all the machine learning methods. For the images that do not contain a vine, they are the same for both sets of machine learning programs.

Figure 2

Example preprocessing output for the VGG16 model using images of the flowering vine.

Figure 3

Confusion Matrix created from VGG16 model using images of the flowering vine



Note: This confusion matrix shows the results of the VGG16 model when using images of the flowering vine. In order to calculate the accuracy, add the left top corner and the right bottom corner divided by 200.

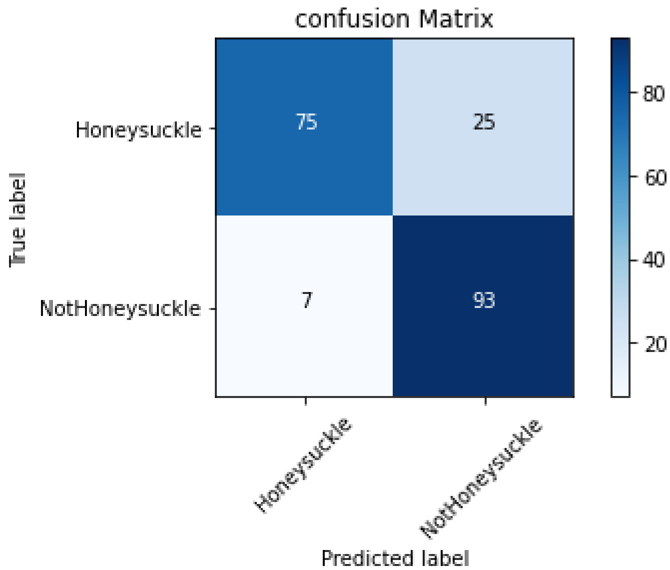
Table 1

The accuracies of each machine learning method with their specified image type

Program Name	CNN with flowering vine images	CNN without flowering vine images	VGG16 with flowering vine images	VGG16 without flowering vine images
Test accuracy %	84	63	94	92.5

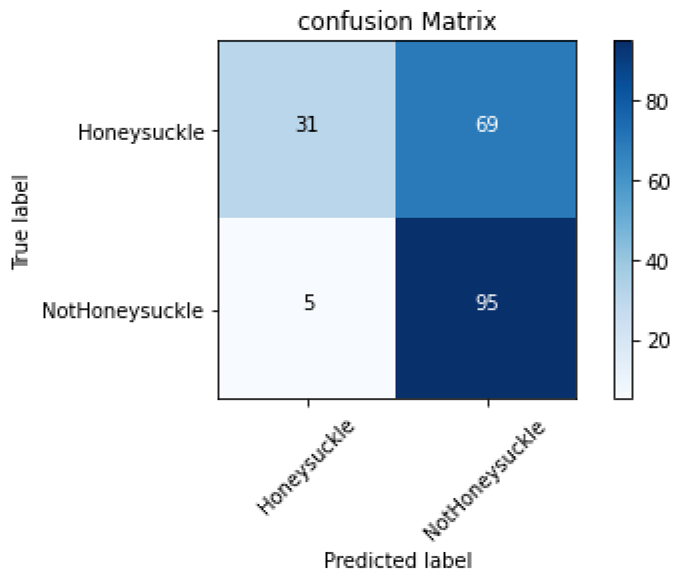
Appendix A

CNN with images of flowering vine confusion matrix



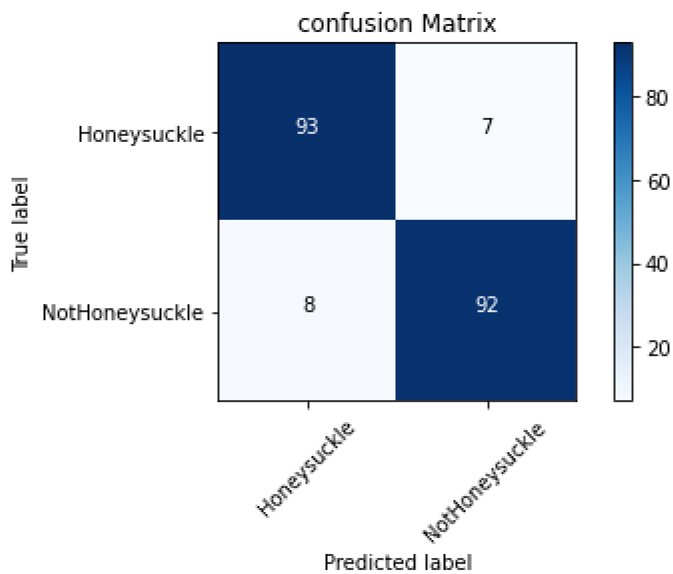
Appendix B

CNN without images of flowering vine confusion matrix



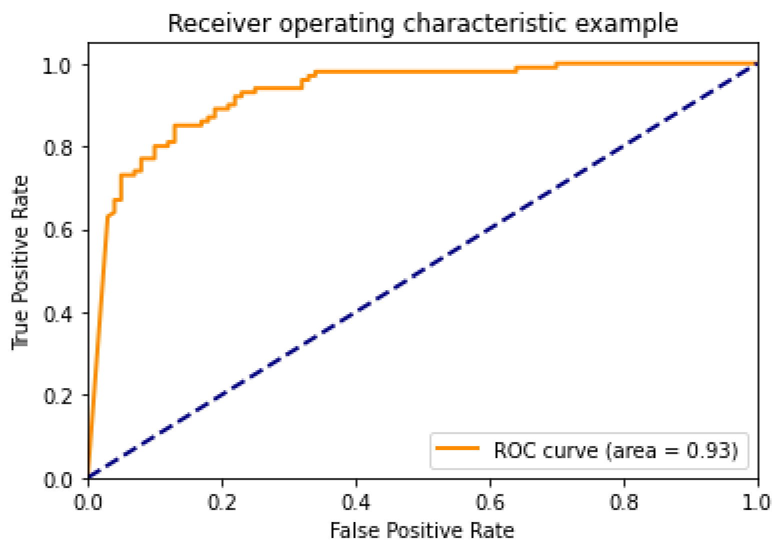
Appendix C

VGG16 without images of flowering vine confusion matrix



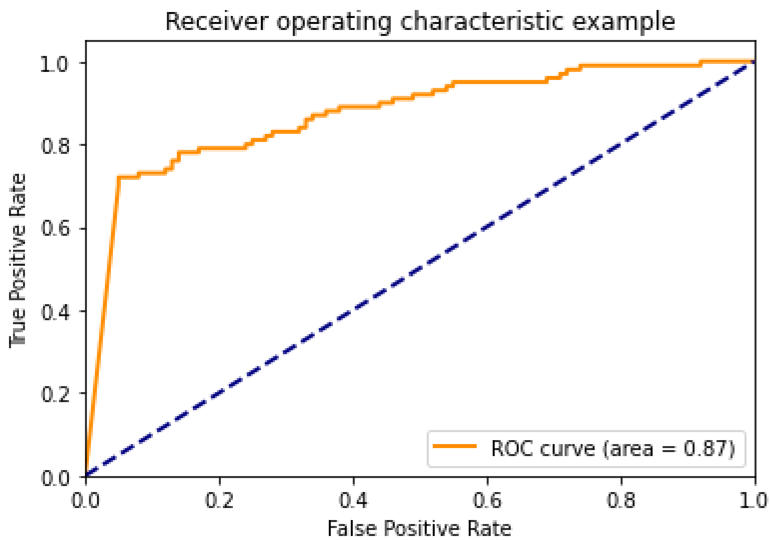
Appendix D

CNN with images of flowering vine ROC graph



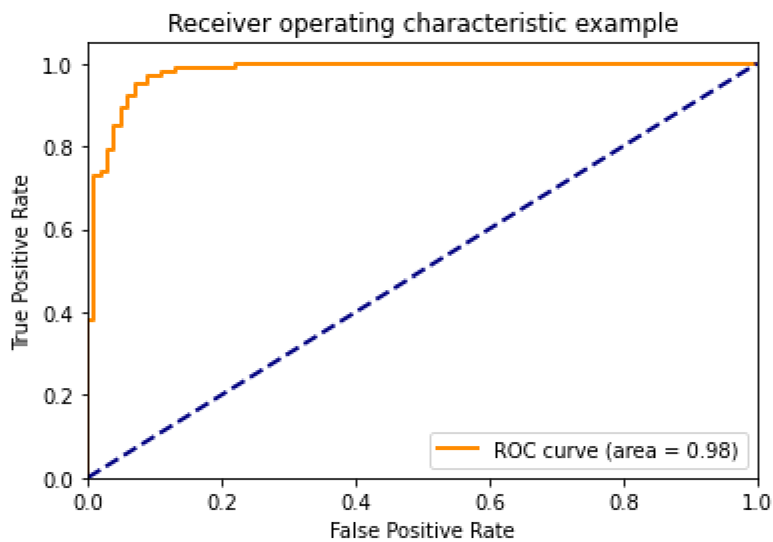
Appendix E

CNN without images of flowering vine ROC graph



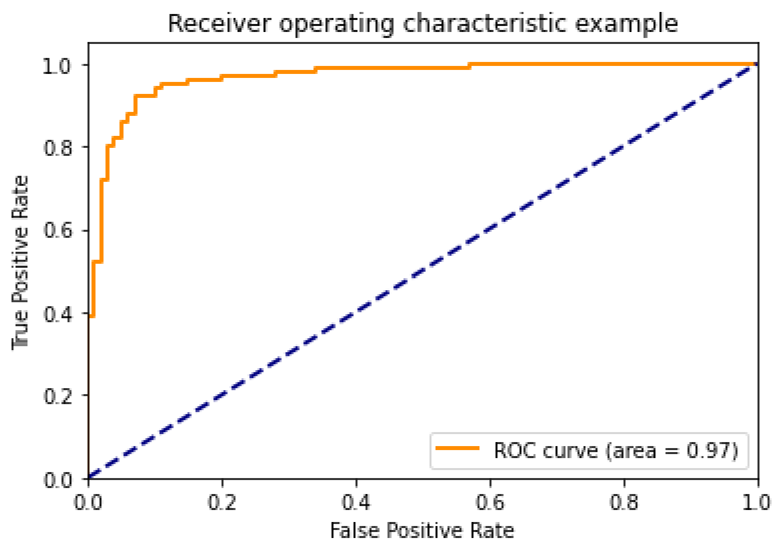
Appendix F

VGG16 with images of flowering vine ROC graph



Appendix G

VGG16 without images of flowering vine ROC graph



Appendix H

Colab link to CNN with images of flowering vine

https://colab.research.google.com/drive/1zdNyMaIR6vDPTDEbjiZfhxbReAXwkMxg?usp=sharing

Appendix I

Colab link to CNN without images of flowering vine

https://colab.research.google.com/drive/1x4i68UW8vG6a3-ViynzbCpKf8ReYMEJp?usp=sharin
g

Appendix J

Colab link to VGG16 with images of flowering vine

https://colab.research.google.com/drive/1p1TDPmsG2Q_U7ZvftEe9Ov4xDOcYIOPt?usp=sharing

Appendix K

Colab link to VGG16 without images of flowering vine

https://colab.research.google.com/drive/1Uv6sMPmwtobJououUP0G6lMPpb2mcDhE?usp=shari
ng