Novel Deception Detection Technique: CasPer Networks and Autoencoders on Facial Thermal Imaging

Rakhmatilla Bakhodirov

Research School of Computer Science
The Australian National University, ACT 2601
u6457384@anu.edu.au

Abstract. CasPer algorithm and feature selection method called AMBER were adopted to detect deception through facial thermal imaging. The algorithms were initially examined in contrast with traditional two layered networks with selected input features by an autoencoder. The results has shown a 7% improvement in detecting whether a person is deceiving or telling the truth. Since the original research paper for the same dataset implements feature selection techniques, this combined algorithm results were compared to the results of the original paper. This novel framework has again demonstrated an improvement in the average accuracy, rising from 71% to 74.2%.

Keywords: Neural Network · CasPer · Deep Learning · Feature Selection · Autoencoders · Deception Detection · Facial Thermal Imaging · Granger Causality · Bioinformatics.

1 Introduction

1.1 Background

Deception detection has always been an important matter throughout the history. Many frameworks exist to detect common patterns that indicate whether a person is deceiving or not. One of these frameworks include analysing facial blood flow changes. Derakhshan, Mikaeili, Nasrabadi, & Gedeon (2018) introduced one such framework to identify these thermal patterns in the cutaneous superficial blood vessels of the face through facial thermal imaging and collected a sample dataset to evaluate this framework. The recorded dataset can be fed into a neural network to predict whether a person is deceiving or telling the truth. This paper will combine both the framework and the technique to achieve the goal: thermal deception detection.

1.2 Motivation

There are many techniques that exist to detect deception, ranging from monitoring of true\false intentions to examination of cues. Yarbrough (2020) states that detectives employ deception detection techniques to "enhance their ability" to distinguish deceit. These deception detection techniques, however, may not always produce reliable results. With an aim to improve the reliability, the CasPer algorithm was employed to detect deception through facial thermal imaging. The CasPer algorithm was chosen from the pool of neural network types because of the following properties (Treadgold & Gedeon, 1998):

· it is a constructive algorithm that automatically finds optimal network structure
· employs Cascade architecture with 3 key regions each with its own learning rate: initially new hidden neurons learn the remaining error with a little interference from other hidden neurons (see Fig. 1).
· uses a variation of RPORP to train the neural network that avoids extreme weights
· does not use weight freezing as opposite to CasCor: old weights can be modified
· smaller network as compared to CasCor
· uses weight decay for generalisation

To further enhance the deception detection accuracy, a computationally efficient wrapper feature selection method called Autoencoder and Model Based Elimination of features (AMBER) was applied (Ramjee & Gamal, 2019). This method makes use of an autoencoder which is a deep learning neural network that is designed to learn a representation of its input by attempting to copy its input to its output (see Fig. 2). As a result, autoencoders are restricted and have to prioritise important features which can be then used to achieve dimensionality reduction to select features with highest information. Meyer, Beutel, & Thiele (2017) report that training an autoencoder on audio events resulted in better clustering of like events in feature space than the baseline.

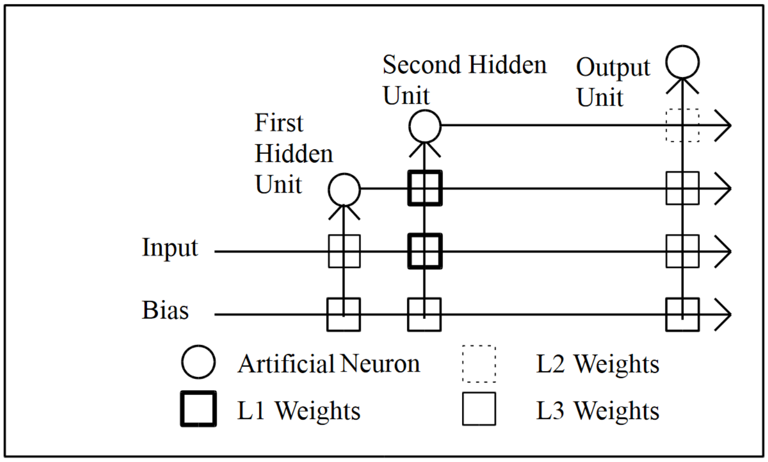

Fig. 1. CasPer Architecture: The second hidden neuron has just been added and vertical lines sum all incoming inputs. Relative values for the learning rates are set by the following criteria L1 >> L2 > L3 (eg. 0.2, 0.005, and 0.001 respectively).

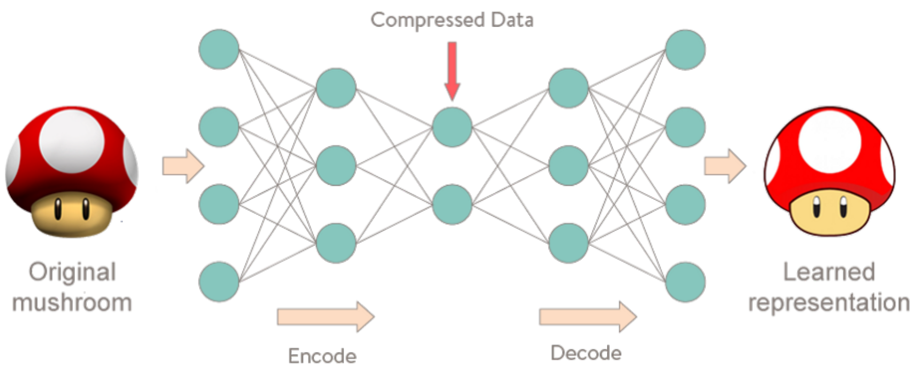

Fig. 2. Autoencoder Architecture Example from `https://medium.com/@curiousily/credit-card-fraud-detection-using-autoencoders`

1.3 Dataset

Derakhshan, Mikaeili, Nasrabadi, & Gedeon (2018) collected a thermal dataset for the purpose of identifying facial patters of the face for 'fight or flight' responses. The dataset included 41 participants of mock crime scenario with equal gender distribution where they performed a fake criminal act of stealing a necklace. For the purpose of the research, five high-sensitivity regions of interest (ROIs) were selected including periorbital (PO), forehead (FH), perinasal (PN), cheek (CK) and chin (CN) as depicted in Fig. 3 (Derakhshan, Mikaeili, Nasrabadi, & Gedeon, 2018). The thermal imaging of superficial blood vessels of the face or simply links between these 5 ROIs where recorded using a modified version of the multivariate Granger Causality (GC) method among each pair of ROIs. There are 20 such pairs which are used as 20 input features of the neural networks outlined in **Method** section of this paper. Each feature is the GC from one region of interest to another.

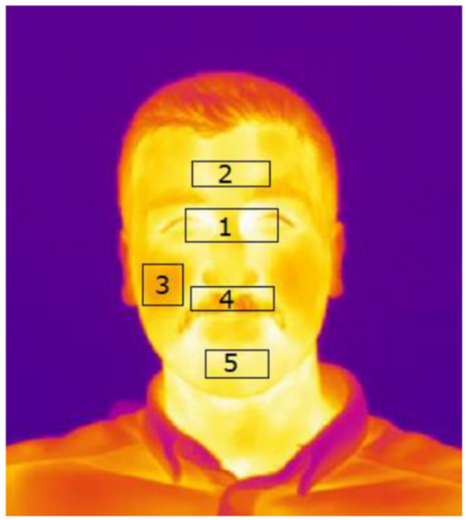

Fig. 3. The ROIs that were segmented based on the anatomy of human facial vasculature including periorbital (1), forehead (2), cheek (3), perinasal (4), and chin (5).

A subset of this dataset was used to evaluate the algorithms that includes 31 participants answering whether they stole the necklace and hence two classes where recorded: 0 (deceptive) and 1 (truthful). Modified dataset was used in favour of the extended dataset for consistency purposes. Moreover, the blood flow between ROIs was assumed to be a time series data. Hence, Derakhshan, Mikaeili, Nasrabadi, & Gedeon (2018) used an extension of the standard Granger Causality (eGC), which shows the causal relationship between two time series data, to determine the relationship between paired facial regions.

Data frequencies were visualised to examine the distribution of the features and the deception detection classification. It can be summarised that the input features generally does not follow the normal distribution with an exception of chin to forehead causality. The class distribution, however, has approximately equal frequencies (see Fig. 4). Besides, the input features are measured to be between 0 and 1, but does not strictly follow this distribution. Even though the eGC indices could have been normalised between 0 and 1 using the min max normalization technique (Cao, Stojkovic, & Obradovic, 2016), it was decided that no input coding is required since the new samples may be fed in range of 0 to 1. Yet, leave-one-out cross validation (LOOCV) was used to provide robustness and reduce bias on limited datasets (Vabalas, Gowen, Poliakoff, & Casson, 2019). Given that this dataset has deception samples first followed by the truth samples, the network can learn to output the second class after seeing a sequence of the same class. Thus, it was decided to try shuffling and splitting the dataset into training and testing sets until an equal number of training and testing samples are generated.

Moreover, a number of investigations were carried out using the proposed model. To start with, the autoencoder neural network was used to determine the number of hidden layers as well as the count of neurons in each layer. Then, sequential feature selection method was employed to rank input features and select top 10 that produce maximum results. Using these selected features, the autoencoder network was again applied to tune the number of hidden layers and the count of neurons per each layer. Once the most efficient architecture was discovered, fully connected neural network as well as CasPer networks were utilized to perform the prediction with both all input features and selected input features. Testing accuracy scores where recorded to evaluate the performances and 5 statistical scores (min, mean, max, median, and standard deviation) were calculated to compare the models.

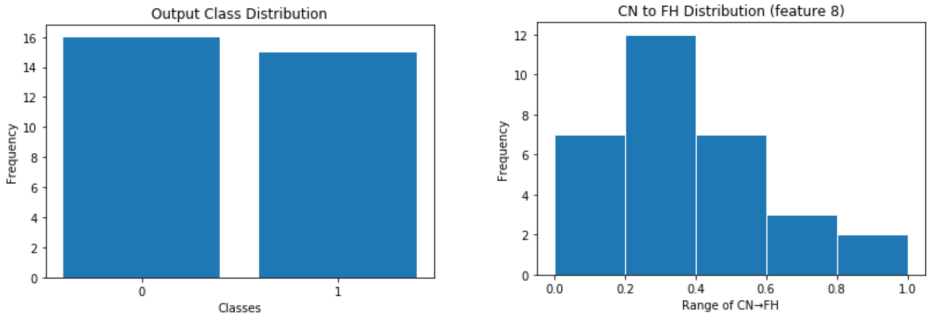

Fig. 4. Frequency distributions.

2 Method

2.1 Autoencoder

Autoencoder and Model Based Elimination of features (AMBER) is a computationally efficient wrapper feature selection method that uses a single ranker model along with autoencoders to perform not only hypertuning of the hidden layers, but also greedy backward elimination of features. Ramjee & Gamal (2019) advise that this model prioritises "the removal of features that are not critical to the classification task, while the autoencoders are used to prioritize the elimination of correlated features". The superior feature selection ability of AMBER on well known datasets, including MNIST and RadioML datasets, were proved to provide higher accuracies as compared to other state-of-the-art techniques (Ramjee & Gamal, 2019). Hence, this technique was employed to conduct investigations on hyperparameters.

The autoencoder neural network was initially used to determine the number of neurons in the first hidden layer and 5 neurons have shown the most efficient result. The same autoencoder was again employed to see whether another hidden layer reduces the MSE loss, but it has not sufficiently reduced the loss (see Fig. 5). Then, sequential feature selection technique was employed to rank input features and select top 10 that produce maximum results. Similarly, the autoencoder was again used to see if adding another hidden layer would improve the performance, but the expectations were not reasonably met as shown in Fig. 6.

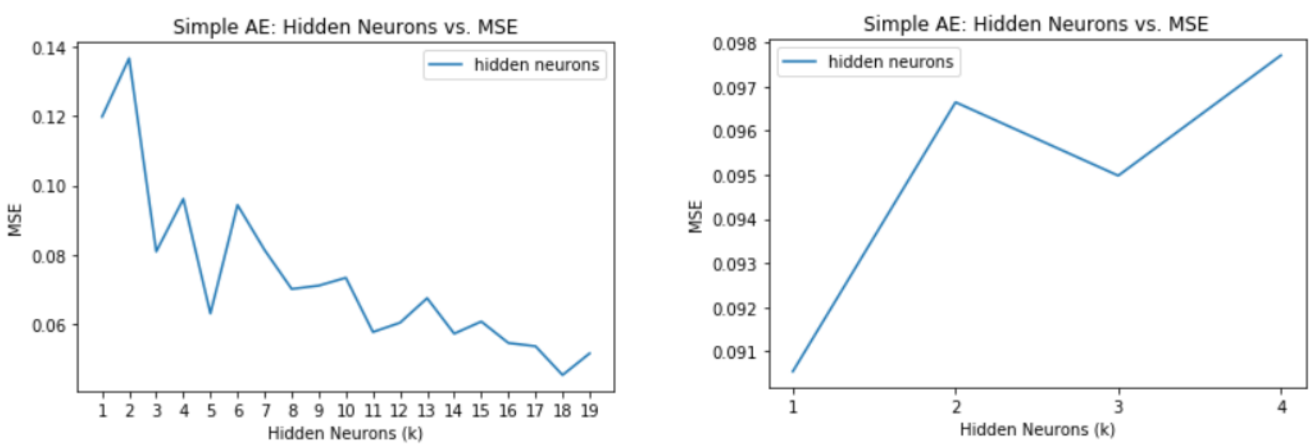

Fig. 5. Hypertuning the number of hidden layers as well as the count of neurons in the first hidden layer using the autoencoder.

2.2 CasPer Algorithm

The CasPer algorithm is a cascade network algorithm employing progressive RPROP that has shown an improvement in both generalization and network size as compared to conventional CasCor algorithms (Treadgold & Gedeon, 1997). Instead of using weight freezing and a correlation measure to install new neurons, CasPer uses a variation of RPROP to train and optimize the whole network (Treadgold & Gedeon, 1998). Each new neuron is linearly connected to all of the previous neurons which makes the neural network fully connected and requires adding the output of each neuron one at a time recursively until it produces the output. This neural network technique also employs 3 different learning rates for 3 separate regions: L1, L2, and L3 as shown in Fig. 1 (Treadgold & Gedeon, 1997).

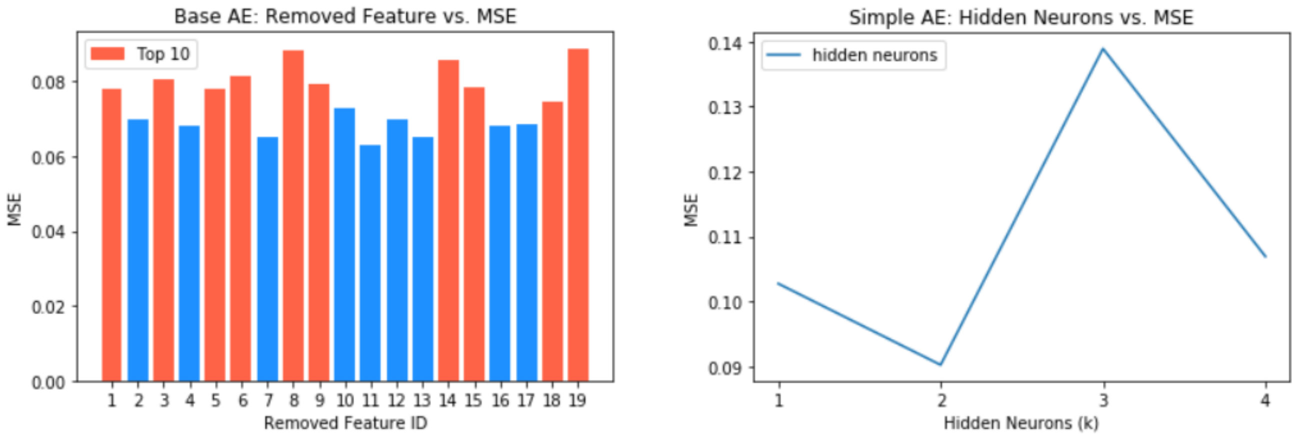

Fig. 6. Sequential feature selection and top 10 selected input features.

2.3 Neural Network Architectures

Fully Connected Neural Network with All Input Features & Selected Features Both neural networks have one hidden layer with 5 neurons and they make use of leave-one-out cross validation (LOOCV) beacuse of having a small dataset.

1. The fully connected neural network on all features:
   · input layer 20 neurons, representing the features of the dataset
   · hidden layer 5 neurons, using LeakyReLU as activation function with a dropout of 0.1 (Nwankpa, Ijomah, Gachagan, & Marshall, 2018)
   · output layer 2 neurons, representing the classes of the dataset
2. The fully connected neural network on selected features defined in Fig. 6:
   · input layer 10 neurons, representing the selected features of the dataset
   · hidden layer 5 neurons, using LeakyReLU as activation function with a dropout of 0.1
   · output layer 2 neurons, representing the classes of the dataset

The networks are trained with Adam as an optimiser, that holds the current state and updates the parameters based on the computed gradients (Kingma & Ba, 2015). The performances are evaluated using cross-entropy. The training is run with 100 epochs. Hyperparameter tuning was conducted that includes:

· Different activations: tanh, sigmoid, LeakyReLU
· Learning rates: 1e-1, 1e-2, 5e-3, 1e-3, 1e-4, 1e-5
· Number of epochs: 5, 10, 50, 100, 200

CasPer Network with All Input Features & Selected Features Two linearly connected neural networks (input to output) are initialised. Then the neural networks are iteratively built one neuron at a time that is fully connected to all existing neurons. CasPer architecture has 3 key regions: L1, L2, and L3; each having separate learning rates: 0.2, 0.005, and 0.001 respectively (Treadgold & Gedeon, 1997).

1. The CasPer neural net on all features has the following layers:
   · input layer 20 neurons, representing the features of the dataset
   · output layer: 2 neurons, representing the classes of the dataset
   · hidden layers: n neurons, using LeakyReLU as activation function with a dropout defined by 3 region learning rates
2. The CasPer neural net on selected features defined in Fig. 6:
   · input layer 10 neurons, representing the selected features of the dataset
   · output layer: 2 neurons, representing the classes of the dataset
   · hidden layers: n neurons, using LeakyReLU as activation function with a dropout defined by 3 region learning rates

The network is trained with Resilient Backpropagation (RPROP) as an optimiser, that holds the current state and updates the parameters based on the computed gradients (Treadgold & Gedeon, 1997). The performances are evaluated using cross-entropy and LOOCV is employed to iterate through different distributions to average the findings. Hyperparameter tuning was conducted that includes:

· Different activations: tanh, sigmoid, LeakyReLU
· Learning rates: 1e-1, 1e-2, 5e-3, 1e-3, 1e-4, 1e-5
· Number of epochs: 5, 10, 50, 100, 200, 500, 1000
· P : 2, 4, 8, 12, 15

P is a hyperparameter that is used for CasPer Network as a time period that serves as a means of early-stopping. If the newly added neuron's error decreases by at least 1% in the time period, then it is added to the network; otherwise the training is stopped completely (Treadgold & Gedeon, 1997). Since the dataset was small, however, 1% threshold was removed altogether to avoid the overfitting.

3 Results and Discussion

Fully Connected Neural Network with All Input Features & Selected Features Table 1 presents summary statistics for both neural networks. The lowest accuracy, median, and the highest accuracies are 0%, 100%, and 100% as expected, whereas the average accuracies are at 54.8% for the neural network with all input features and 67.7% for the neural network with selected features respectively. Even though the fully connected neural network with selected features demonstrated high accuracy score, the standard deviation scores for these networks are 49.8% and 46.8% respectively. This high variances can be explained simply by the fact that there is a single sample to test the neural network on which it produces either correct or incorrect result. Hence, there is approximately 50% chance for the network to detect deception. While leave-one-out cross-validation is approximately unbiased, it tends to have a high variance. Very different estimates could have been produced if the estimate was repeated with different initial samples of data from the same distribution.

Table 1. Statistics to evaluate the performance of Fully Connected Neural Networks.

Statistic Scores in LOOCV	Score for All Features	Score for Selected Features
Lowest accuracy	0.0%	0.0%
Average accuracy	54.84%	67.74%
Highest accuracy	100.0%	100.0%
Median accuracy	100.0%	100.0%
Standard deviation	49.77%	46.75%

CasPer Network with All Input Features & Selected Features Similarly to the fully connected neural networks analysis, table 2 depicts summary statistics for both neural networks. The lowest accuracy, median, and the highest accuracies remain the same, whereas the average accuracy and the standard deviation show different statistics on CasPer neural network with selected input features. Specifically, the average accuracy increases by almost 7% (74.2%), while the standard deviation drops by 3% (43.8%). Surprisingly, the fully connected neural network and CasPer networks with all features show the same accuracies for all summary statistics. This can be because of learning the same features even though having different architectures.

Table 2. Statistics to evaluate the performance of CasPer Networks.

Statistic Scores in LOOCV	Score for All Features	Score for Selected Features
Lowest accuracy	0.0%	0.0%
Average accuracy	54.84%	74.19%
Highest accuracy	100.0%	100.0%
Median accuracy	100.0%	100.0%
Standard deviation	49.77%	43.76%

Original Research Paper Comparison Given that Derakhshan, Mikaeili, Nasrabadi, & Gedeon (2018) researched on feature selection models, it cannot be closely analysed against this paper. However, Bakhodirov (2021) has modeled the CasPer network on selected features outlined by Derakhshan, Mikaeili, Nasrabadi, & Gedeon (2018) and the summary statistics can be found in table 3.

It can be seen that the efficient wrapper feature selection method, AMBER, indicates better performance as compared to the method outlined in the original research paper by almost 4% average accuracy. The standard deviation has also improved from 45.4% to 43.8% respectively.

Table 3. Statistics to evaluate the performance of CasPer Network on selected features outlined in the original research paper.

Statistic Scores in LOOCV	Score
Lowest accuracy	0.0%
Average accuracy	70.97%
Highest accuracy	100.0%
Median accuracy	100.0%
Standard deviation	45.39%

4 Conclusion and Future Work

An improvement in detecting deception through thermal imaging was made by employing CasPer algorithm and AMBER technique as compared to traditional two layered networks. The results has shown a 7% improvement in detecting whether a person is deceiving or telling the truth. Besides, CasPer networks with selected input features have added another 4% average accuracy to deception detection performance as compared to the original research paper findings, increasing the testing accuracy from 71% to 74.2%. Both CasPer algorithm and feature selection have, therefore, increased deception detection accuracy. Given that leave-one-out cross validation technique has a single sample to test the accuracy of the algorithm, the accuracy score produced was either 0% or 100%. Leave-two-out cross validation could be adopted in the future to improve the accuracy of detecting whether an individual is deceiving or telling the truth. Moreover, the problem of having small dataset could be overcome through data augmentation to make the most of the data.

References

1. Bakhodirov, R. (2021). Truth or Lie? CasPer Networks in Thermal Deception Detection. *4th ABCs ANU Bio-inspired Computing Conference.* Retrieved from `https://easychair.org/conferences/submission_download?submission=5458853;upload=113404;a=26515364`
2. Cao, X. H., Stojkovic, I., & Obradovic, Z. (2016). A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics, 17*(1). Retrieved from `https://dx.doi.org/10.1186/s12859-016-1236-x`
3. Derakhshan, A., Mikaeili, M., Nasrabadi, A. M., & Gedeon, T. (2018). Network physiology of 'fight or flight' response in facial superficial blood vessels. *Physiological Measurement, 40*(2019), 014002. Retrieved from `https://doi.org/10.1088/1361-6579/aaf089`
4. Jacobs, G. D. (2001). The Physiology of Mind-Body Interactions: The Stress Response and the Relaxation Response. *The Journal of Alternative and Complementary Medicine., 7*, 83-92. Retrieved from `https://doi.org/10.1089/107555301753393841`
5. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations.* San Diego. Retrieved from ResearchGate: `https://www.researchgate.net/publication/269935079_Adam_A_Method_for_Stochastic_Optimization`
6. Meyer, M., Beutel, J., & Thiele, L. (2017). Unsupervised Feature Learning for Audio Analysis. *5th International Conference on Learning Representations.* Retrieved from `https://arxiv.org/pdf/1712.03835.pdf`
7. Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. Retrieved from ResearchGate: `https://www.researchgate.net/publication/328826136_Activation_Functions_Comparison_of_trends_in_Practice_and_Research_for_Deep_Learning`
8. Ramjee, S., & Gamal, A. E. (2019). Efficient Wrapper Feature Selection using Autoencoder and Model Based Elimination. *ArXiv.* Retrieved from `https://arxiv.org/pdf/1905.11592.pdf`
9. Sonkusare, S., Ahmedt-Aristizabal, D., Aburn, M. J., Nguyen, V. T., Pang, T., Frydman, S., . . . Guo, C. C. (2019). Detecting changes in facial temperature induced by a sudden auditory stimulus based on deep learning-assisted face tracking. *Scientific Reports (Nature Publisher Group), 9*(1), 1-12. Retrieved from `https://doi.org/10.1038/s41598-019-41172-7`
10. Treadgold, N. K., & Gedeon, T. D. (1997). A Cascade Network Algorithm Employing Progressive RPROP. *International Work-Conference on Artificial Neural Networks, 1240.* Retrieved from `https://doi-org.virtual.anu.edu.au/10.1007/BFb0032532`
11. Treadgold, N., & Gedeon, T. D. (1998). Exploring Architecture Variations in Constructive Cascade Networks. *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227), 1*, 343-348. Retrieved from `https://www.doi.org/10.1109/IJCNN.1998.682289`
12. Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE, 14*(11), e0224365. Retrieved from `https://doi.org/10.1371/journal.pone.0224365`
13. Yarbrough, J. R. (2020). The Science of Deception Detection: A Literature and Policy Review on Police Ability to Detect Lies. *The Science of Deception Detection: A Literature and Policy Review , 3*(2), 47-66. Retrieved from `https://www.uhd.edu/academics/public-service/jcjl/Documents/Volume3Issue2Yarbrough.pdf`