The effect of Data Pre-processing technique on depression level prediction

Tao Xu[1]

[1] Research School of Computer Science, Australian National University, u5527268@anu.edu.au

Abstract. Depression level prediction can be achieved by using a neural network with physiological responses data. The data pre-processing technique plays a vital role in building the neural network model. This report demonstrates using normalization data processing method to improve the neural network performance and generate a meaningful prediction of the depression level prediction. Moreover, physiological responses data is time-series data and Long Short-Term Memory (LSTM) RNN was used to predict the depression level as LSTM has an excellent performance in time-series data. Also, the final results compared with the original dataset paper, which approved the importance of data pre-processing during the neural network building process.

Keywords: depression, neural network, data, LSTM

1 Introduction

Depression is a mental health disorder and it is characterized by persistent emotional depression or loss of interest in activities, which affects the daily life a lot. The possible causes may come from the physical, psychological, and social environment. More and more studies have shown these factors may cause changes in brain function as well as the activity of the certain neural circuits in the brain. Depression is defined by persistent emotional sadness or loss of interest, which results in a series of behavioural and physical symptoms. These symptoms may include changes in sleep, appetite, energy, attention, daily behaviour, or self-esteem. People with depression may also have suicidal thoughts. There are three main kinds of depression which are Major depressive disorder, Persistent Depressive Disorder, Seasonal Affective Disorder. Major depressive disorder (MDD) is the most serious depression symptoms that demonstrate in three aspects: mood cognition and somatic functional mood. Cognitively, patients tend to see the negative side of things and are surrounded by the feeling of emptiness and wothlessness[1]. Persistent Depressive Disorder symptoms are similar to the MDD, but are less severe and last much longer than MDD[2]. Comparing with MDD, Seasonal Affective Disorder symptoms are similar to or less severe than MDD, and it is classified as a subtype of depression[3]. The main treatment method of depression usually includes medication, talk therapy or a combination of two. Some treatment may promote changes in brain function that is related to depression return to normal[4]. Diagnosing depression is fundamentally clinical, which is done by a clinical interview and mental status examination[5]. Moreover, the diagnosis of depression should be based on medical history, clinical, symptoms, physical examination. Normally, the diagnostic standards include ICD-10 and DSM-IV[6]. ICD-10 is used frequently in diagnosing depression, which refers to the first episode of depression and recurrent depression but excluding bipolar depression[7].

As the diagnosing depression always need a clinical interview or patients self-reported, which is time-consuming. Consequently, some deep learning diagnosing methods were developed. The deep learning method can detect the depression of social media user by using neutral language processing[8]. Also, the neural network can facilitate to diagnose the adult depression level[9]. The neural network has the potential to diagnose depression with high accuracy. I think the neural network provides a potential way to diagnosis depression, which could make a great contribution to the human community. Therefore, I choose the depression data set that is provided by the paper "Detecting emotional reaction to videos of depression"[10]. They collect the physiological responses data and use them to detect the level (minimal, mild, moderate, severe) of the depression. The physiological responses include galvanic skin response (GSR), skin temperature (ST) and pupillary dilation (PD). There are 85 features from the three physiological signals which include 23 galvanic skin response (GSR) signal data, 39 pupillary dilations (PD) data and 23 skin temperature data. GSR is a kind of electrodermal activity responses, which is used to measure the current flow of the skin and the current changes with the sweat level of the skin[11]. The GSR has two main components which are tonic component and phasic component. The tonic component was used to detect the activity of the perspiratory glands which related to the body external temperature. The phasic component is linearly related to the intensity of arousal in mind state[12]. Vasodilatation causes skin temperature changes. The stress and fear emotions are negatively correlated with the skin temperature[13]. Pupillary dilations demonstrated the changes in the mental activities and the pupil size will increase after positively and negatively stimuli[14]. In "Detecting emotional reaction to videos of depression" paper, they use a genetic algorithm to select the data features. The genetic algorithm features selection is a wrapper method. The algorithm uses the recognition rate of the support vector machine classifier as the judgment basis for feature selection. In the algorithm, the selected features are initialized with 0 and 1 binary strings. Since 0 and 1 occur with equal probability. The expectation of the optimal number of features is half of the original number of features. To further reduing the number of features, it can make the binary number 0 and 1 appear with unequal probabilities[15]. This network is a classifier that classifies the participants into different classes of depression. For training this classifier model, the physiological responses data was treated as input and the depression class type worked as output. Finally, they found the GSR, ST and PD are realted to the depression levels and they believe using neural networks can facilitate the depression diagnosis.

The other paper "Decrypting neural network data: a GIS case study" provided a method to process different types of data and improve the performance of the neural network on the data[16]. It focuses on how to pre-process the input and output training data and try to make them suitable to apply to the neural network to achieve high performance. This method was used my work to analyze the physiological responses data and then compare neural network performance difference with the original result. In this paper, there are many different types of data and the author processed them in different ways. The data includes aspect, altitude, topographic position, slope degree, rainfall, temperature and Landsat tm bands 1 to 7. These data have different range and different types (nominal, ordinal, numerical). The aspect and slope degree data are circular value data and the altitude, rainfall, temperature are numerical data with different ranges. The paper processed these data in different ways. The deep learning neural network requires a number variable for input and output. Hence, all the data was transferred to numerical data[16]. In this report, the paper "Decrypting neural network data: a GIS case study" method was used to process the physiological data and the use these data to train the network to check if the method improved the performance of the neural network. As above mentioned, this report aims to predict the depression level by using the GSR, Pupil, Skin-temperature data that was processed by using "Decrypting neural network data: a GIS case study" paper's method.

2 Method

As mentioned above, depression diagnosing plays a vital role in depression treatment. Using physiological responses data to do depression level prediction is a direction with quite a development potentiality. Consequently, the raw physiological responses data should be processed well before using to build the neural network. The "Detecting emotional reaction to videos of depression" paper provided 3 excel file to store the physiological responses data, which includes galvanic skin response (GSR), skin temperature (ST) and pupillary dilation (PD) data. There are 23 features were generated from normalized and filtered GSR data. They are normalized minimum value, normalized maximum value, normalized mean value, normalized standard deviation value, normalized variance, normalized root mean square, filtered minimum value, filtered maximum value, filtered mean value, filtered standard deviation value, filtered variance, filtered root mean square, means of the absolute values of the first difference, means of the absolute values of the second difference, means of the absolute values of the first difference, means of the absolute values of the second difference, normalized number of peaks, filtered number of peaks, normalized numbers of SCR occurrences for VLP, LP, normalized amplitudes of SCR occurrences for VLP, LP, ratio of SCR occurrences in VLP to occurrences in LP. The pupillary dilation and skin temperature data have the same type of data as GSR. Some of the features can not be used to train the neural network because they may redundant data, which will affect the classification performance of the neural network. These data need to be processed by using different methods such as normalization, absolute value and fixing missing value. Therefore, it needs to select a proper method and process the data before training the neural network. As above mentioned, these physiological responses data was selected by using the generic algorithm. In this report, the paper "Decrypting neural network data: a GIS case study" data processing method was used to process the physiological responses data.

According to the investigation of three datasets, all the data is numerical, which is same as the altitude, rainfall, temperature of GIS data[16]. The paper process these numerical data by using the normalized method and all the numerical data was put into range 0-1. However, in the GSR dataset, the number of normalized GSR peaks, the number of filtered GSR peaks, the value of VLP occurrences and mean LP amplitudes are not in the range. The value does not in the range 0-1 so the normalize method was applied to that data. To transfer data to range 0-1, the formula (1) was used to arrange the data to range 0-1. The classifier output has been encoded as 0,1,2,3, which represents minimal, mild, moderate and severe. The output can be employed on network training. Furthermore, there are some features include negative value, which is not in range 0-1. Therefore, these data were processed by using the absolute value function.

$$x = \frac{\mathrm{Xi} - \min()}{\max() - \min()} \tag{1}$$

For comparing the result with the "Detecting emotional reaction to videos of depression" paper, the variable-controlling approach was used in this report. The data processing method comes from the "Decrypting neural network data: a GIS case study" paper as it focuses on data processing and did not mention neural network in detail. It only mentioned 12 hidden neurons were used and also used different epoch numbers to check the train and test accuracy. Therefore, the neural network parameters use a hidden layer with size 50, four output neurons, adam optimizer[17] and cross-entropy loss function, which comes from "Detecting emotional reaction to videos of depression" paper[10]. The adam optimizer can be considered as a combination of RMSprop and Momentum and the exponential moving average is used for the firstorder momentum in Adam, which can guarantee iteration to be relatively smooth. Therefore, it was used for the network optimizer. The loss function is the cross-entropy loss function. Cross-entropy was always used in the classification problem, which described the distance between two probability distributions. The smaller the cross-entropy is, and the closer they are. In this report, we are facing a classification problem so it was used. The "Decrypting neural network data: a GIS case study" paper technique is the backbone of the depression level prediction. The evaluation method also follows the paper method, using precision, recall and F1 score, which is introduced by the "Detecting emotional reaction to videos of depression" paper[10]. The precision is the proportion of predicting depression level of people who has depression. The recall is the percentage of a depressed person who is predicted correctly with the depression label. F1 calculation uses the formula (2).

Moreover, for the model design, all physiological responses data were used in the training dataset, which includes pupil, GSR and skin temperature. The reason for using all the data is the dataset paper using all features to train the network and predict the depression level. If we only use partly data, we cannot ensure if the data processing method affected the performance of the network. Consequently, these three datasets should be combined. There are 85 features and 1 label and the label includes 4 classes. Hence, the model input layer has 85 inputs and 4 outputs in the output layer and the hidden layer size is 50, which is derived from the dataset paper and they found 50 is optimal for their task [10]. The number of epoch is 500. In addition, there are some details of the technique. The technique paper processed the numerical data with the normalization method. As mentioned above, there are some values not in range 0-1. The maximum function and minimum function were applied on the off-limits column to find the max value and min value. And each value of this column subtract the min value and then divide the difference between max value and min value. Besides, there are some negative values in the dataset so all of them need to change to the positive value. The absolute value function was used for that. After training the networks, the confusion matrix was generated and it also was used for calculating precision, recall and F1 score. Furthermore, the un-preprocessed data also was used to build the network. It aims to compare with the pre-processed data result. The overall accuracy is calculated by $\frac{\text{correct prediction}}{\text{total data size}} * 100\%$.

$$F1 = \frac{2*precision*recall}{(\text{precision}+\text{recall})} \tag{2}$$

RNN is one of the most commonly used models to deal with time-series problems. RNN has an excellent performance in time-series data as RNN at (t) time slice takes (t-1) time slice hidden nodes as input. The reason for this is that the information of the previous time slice is also used to calculate the content of the current time slice, while the output of the hidden nodes of the traditional model only depends on the input characteristics of the current time slice. However, RNN has the problem of Long Term Dependencies so LSTM was raised to tackle this problem [18]. LSTM is a neural network capable of memorizing short and long term information, which is demonstrated in 1997 [19]. While the LSTM can solve the long term dependence problem as it introduces the gate mechanism to control the flow and loss of features [20]. In this report, the hidden size of the hidden layer is 180. The output size is 4 as there are 4 labels. The optimiser is Adam and the loss function is cross-entropy. The normalized data and non-normalized data were used to train and test the network. Both data were also separated into train and test data. There are 192 time-series data, and each one was collected every ¼ second. The first 100 data was used as train data and left 92 data was used as test data.

3 Results and Discussion

Based on the technique introduction, the number of input neurons is 85. The number of hidden neurons is 50. The number of output neurons is 4 and the number of the epoch is 500. With these parameters, the un-preprocessed data was used to train the network. The training loss is 0.2664 at epoch 400 and the training accuracy is 94.16%. Moreover, the training loss is 0.2081 and the accuracy is 96.75% at epoch 500. The trained network also was used to predict the test data and the accuracy is 39.47% Then the pre-processed data was used to train the network. The training loss is 0.1547 and the accuracy is 99.32% at epoch 400. When the epoch reaches 500, the training loss is 0.1070 and the accuracy is 100%. The pre-processed data trained network was used to predict test dataset, the accuracy is 43.48%. According to the result, the model is overfitting[21] due to the small dataset. When the model is overfitting, dropout function is a solution to tackle this problem[22]. The Pytorch has built-in dropout function and the 0.5 was set for the dropout function. 0.5 refers to the 50% neurons are closed or discarded randomly. After the dropout function was added, the un-preprocessed data training loss is 0.2333 with the accuracy of 98.70% at epoch 400 and the loss decreases to 0.1862 at epoch 500. The test accuracy increase to 42.11%. The pre-processed data training loss is 0.2509 and the accuracy is 96.60% at the epoch 400. At the epoch 500, the loss is 0.1917 and accuracy is 97.96%. The test accuracy increase to 48.89%. The confusion matrix was used to calculate the precision, recall and F1 score. Also, early stopping is a normal method to solve the overfitting problem [23]. For the early stopping, the original data set was divided into training data set, validation data set and test dataset. After training the network, the validation data set was used to calculate the error rates of the network for each cycle. If the error is worse than the result of the last training, stop training. In this report, the validation data is the same as the test data due to the small size of the whole dataset. And then using the parameters of the last iteration results as the final parameters of the model. The un-preprocessed data training accuracy of the network increased to 80.50% after using the early stopping and dropout function and the pre-processed data training model accuracy is 90.70%. And the precision, recall and F1 score increased. The result of pre-processed data training neural networks was shown in table 1, which includes without dropout, with dropout and dropout and early stopping. The table2 shows the un-preprocessed data result of the test result with different methods. The dataset paper represented NN and GA+NN model results. The NN uses all the features derived from the physiological signal to train the network[10]. It got overall accuracy 0.88. The GA+NN uses GA to select features and then to train the network. It gets overall accuracy 0.92. The technique paper provided data processing technique to pre-process the data before using. It normalized all the numerical data in the same range. This report uses pre-processed data to train the network without features selection. In the beginning, the accuracy is around 40%~50% due to the overfitting problem. However, the overfitting can be tackled by dropout and early stopping. Finally, the network with the pre-processed data can achieve 90.70% accuracy, which is only 1.3% lower than the model with GA feature selection, at a rate of 92%. However, the implementation of the genetic algorithm is relatively complex. Firstly, the problem needs to be encoded, and then need to decode the problem after the optimal solution is found. Secondly, three operators also have many parameters, such as crossover rate and mutation rate. The selection of these parameters seriously affects the quality of the solution, while the selection of these parameters is mostly based on experience[24]. The dropout and early stopping functions have been build in Pytorch, and it is easy to implement and achieve a high-performance network. The technique paper provided a vital input and output data pre-processing method, which is an essential step for training a neural network.

TABLE 1 PERFORMANCE MEASURE WITH DIFFERENT METHODS(pre-processed data)

	Without dropout			**With dropout**			**Dropout and early stopping**		
Depression Level	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
None	0.5	0.63	0.56	0.5	0.63	0.56	0.92	0.92	0.92
Mild	0.33	0.45	0.38	0.33	0.45	0.38	1.0	0.875	0.93
Moderate	0.33	0.38	0.35	0.33	0.38	0.35	0.92	0.86	0.89
Severe	0.4	0.2	0.27	0.4	0.2	0.27	0.8	1.0	0.9
Overall accuracy	43.48%			48.89%			90.70%		

Table 2 PERFORMANCE MEASURE WITH DIFFERENT METHODS(un-processed data)

	Without dropout			**With dropout**			**Dropout and early stopping**		
Depression Level	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
None	0.40	0.52	0.50	0.32	0.42	0.47	0.70	0.82	0.82
Mild	0.21	0.42	0.24	0.29	0.32	0.34	0.84	0.75	0.85
Moderate	0.22	0.35	0.33	0.30	0.27	0.29	0.62	0.86	0.79
Severe	0.38	0.2	0.18	0.4	0.2	0.16	0.7	0.82	0.74
Overall accuracy	39.47%			42.11%			80.50%		

For the RNN LSTM with normalized data, the training loss is 0.409 and the accuracy is 89% at epoch 180. The training loss is 0.3779 and the accuracy is 90% at epoch 200. The trained LSTM was used to predict the test data. The accuracy of the testing is 31.52%. And Table 3 shows the normalized data precision, recall and F1 score.

TABLE 3 LSTM (pre-processed data)

Depression Level	**Precision**	**Recall**	**F1 score**
None	0.33	0.55	0.415
Mild	0.4375	0.2916	0.35
Moderate	0.2608	0.25	0.255
Severe	0.25	0.208	0.2272
Overall accuracy	31.52%		

For the RNN LSTM with non-normalized data, the training loss is 0.1812 and the accuracy is 98% at epoch 180. The training loss is 0.1547 and the accuracy is 99% at epoch 200. The trained LSTM was used to predict the test data. The accuracy of the testing is 30.43%. And Table 4 shows the non-normalized data precision, recall and F1 score.

Table 4 LSTM (unprocessed data)

Depression Level	**Precision**	**Recall**	**F1 score**
None	0.3	0.3	0.3
Mild	0.333	0.2917	0.311
Moderate	0.2647	0.375	0.310
Severe	0.3529	0.25	0.2927
Overall accuracy	30.43%		

According to the normalized and non-normalized result, the training accuracy is very high but the test accuracy is low, which means the overfitting occurred during the training process. The dropout will be applied first. The non-normalized testing accuracy is 32.61% that increased 2% after adding dropout. The normalized data testing accuracy is 35.87% after adding the dropout. Therefore, early stopping need to apply on LSTM again. Therefore, it needs test data, training data and valid data. The valid data was picked up from the 80% of the dataset. After the early stopping was used in the LSTM, the normalized data test accuracy is 89.13% and the highest training accuracy is 90%. The highest non-normalized data training accuracy is 100% and the testing accuracy is 83.70%. Figure 1a) shows the normalized data training loss per epoch and accuracy per epoch. Figure 1 b) shows the non-normalized data training loss per epoch and accuracy per epoch.

a)

```
Epoch [1/200] Loss: 1.3909  Accuracy: 28.00 %
Epoch [11/200] Loss: 1.3695  Accuracy: 31.00 %
Epoch [21/200] Loss: 1.3200  Accuracy: 46.00 %
Epoch [31/200] Loss: 1.2559  Accuracy: 46.00 %
Epoch [41/200] Loss: 1.1862  Accuracy: 53.00 %
Epoch [51/200] Loss: 1.1015  Accuracy: 55.00 %
Epoch [61/200] Loss: 1.0234  Accuracy: 59.00 %
Epoch [71/200] Loss: 0.9399  Accuracy: 63.00 %
Epoch [81/200] Loss: 0.8961  Accuracy: 63.00 %
Epoch [91/200] Loss: 0.8364  Accuracy: 70.00 %
Epoch [101/200] Loss: 0.7759  Accuracy: 73.00 %
Epoch [111/200] Loss: 0.7237  Accuracy: 78.00 %
Epoch [121/200] Loss: 0.6678  Accuracy: 76.00 %
Epoch [131/200] Loss: 0.6127  Accuracy: 79.00 %
Epoch [141/200] Loss: 0.5788  Accuracy: 84.00 %
Epoch [151/200] Loss: 0.5210  Accuracy: 86.00 %
Epoch [161/200] Loss: 0.4730  Accuracy: 86.00 %
Epoch [171/200] Loss: 0.4233  Accuracy: 86.00 %
Epoch [181/200] Loss: 0.4076  Accuracy: 88.00 %
Epoch [191/200] Loss: 0.3560  Accuracy: 90.00 %
```

b)

```
Epoch [1/200] Loss: 1.4010  Accuracy: 24.00 %
Epoch [11/200] Loss: 1.2795  Accuracy: 38.00 %
Epoch [21/200] Loss: 1.1890  Accuracy: 49.00 %
Epoch [31/200] Loss: 1.0556  Accuracy: 53.00 %
Epoch [41/200] Loss: 0.9370  Accuracy: 64.00 %
Epoch [51/200] Loss: 0.8434  Accuracy: 74.00 %
Epoch [61/200] Loss: 0.7821  Accuracy: 71.00 %
Epoch [71/200] Loss: 0.7092  Accuracy: 74.00 %
Epoch [81/200] Loss: 0.6252  Accuracy: 80.00 %
Epoch [91/200] Loss: 0.5524  Accuracy: 81.00 %
Epoch [101/200] Loss: 0.4825  Accuracy: 84.00 %
Epoch [111/200] Loss: 0.4286  Accuracy: 89.00 %
Epoch [121/200] Loss: 0.3623  Accuracy: 90.00 %
Epoch [131/200] Loss: 0.3090  Accuracy: 95.00 %
Epoch [141/200] Loss: 0.2669  Accuracy: 94.00 %
Epoch [151/200] Loss: 0.2155  Accuracy: 97.00 %
Epoch [161/200] Loss: 0.1823  Accuracy: 97.00 %
Epoch [171/200] Loss: 0.1519  Accuracy: 99.00 %
Epoch [181/200] Loss: 0.1274  Accuracy: 100.00 %
Epoch [191/200] Loss: 0.1077  Accuracy: 100.00 %
```

Figure 1 loss per epoch and accuracy per epoch.

4 Conclusion and Future Work

Depression is a common disease and more than 264 million people are suffering that[25]. Therefore, it is important to diagnose depression accurately. The neural network provided a powerful approach to predict the depression level. Different kinds of the physical sensor were used to collect raw physiological responses data of patients. These data were necessary to pre-processed before using them to train the neural network model. Data pre-processing is a vital step before training a neural network. Most people spend 80% time on processing data and only use 20% time on building a network or writing algorithm. The quality of data directly determines the prediction and generalization ability of the model. It includes many factors, including accuracy, completeness, consistency, timeliness, credibility, and interpretability. If the data quality is not good, the best model can not work well. In the real world, the data that we get may contain a large number of missing values and may contain a lot of noise or may input errors. There are some methods to preprocess the data, such as data cleaning, data normalization and data transfer. Most of the raw physiological responses data of patients are numerical data. For numerical dataset, data normalization is a vital step for data processing. Normalization makes all the data in the same range. It can avoid training would worse as it would be dominated by the inputs with the highest value. Also, avoid training would be worse as the model would consider that the higher values were more important. With the pre-processed data, the trained model has a better performance than the model which use raw data. According to the result of this paper, the normalized data has a better performance in normal neural networks and RNN LSTM. Pre-processed data is better than un-preprocessed data for training a model. In the future, the better data processing technique should be developed and use different kinds of data pre-processing skills to improve the network performance on data.

Reference

1. Barlow, D. and V. Durand, *Abnormal psychology: An integrative approach*. Nelson Education(2011).
2. Gater, R., et al., *Persistent depressive disorders and social stress in people of Pakistani origin and white Europeans in UK.* **44**(3): p. 198(2009).
3. Lurie, S.J., et al., *Seasonal affective disorder.* **74**(9): p. 1521-1524(2006).
4. Beck, A.T. and B.A. Alford, *Depression: Causes and treatment*. University of Pennsylvania Press(2009).
5. Goldman, L.S., et al., *Awareness, diagnosis, and treatment of depression.* **14**(9): p. 569-580(1999).
6. Almeida, O.P. and S.A.J.I.j.o.g.p. Almeida, *Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV.* **14**(10): p. 858-865(1999).
7. Löwe, B., et al., *Diagnosing ICD-10 depressive episodes: superior criterion validity of the Patient Health Questionnaire.* **73**(6): p. 386-390(2004).
8. Orabi, A.H., et al. *Deep learning for depression detection of twitter users*. in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* of Conference. (Year).
9. Chattopadhyay, S., et al., *Neural network approaches to grade adult depression.* **36**(5): p. 2803-2815(2012).
10. Zhu, X., et al. *Detecting emotional reactions to videos of depression*. in *IEEE International Conference on Intelligent Engineering Systems* of Conference. (Year).
11. Stern, R.M., et al., *Psychophysiological recording*. Oxford University Press, USA(2001).
12. Kim, J., E.J.I.t.o.p.a. André, and m. intelligence, *Emotion recognition based on physiological changes in music listening.* **30**(12): p. 2067-2083(2008).
13. Sharma, N. and T. Gedeon. *Modeling stress recognition in typical virtual environments*. in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops* of Conference.: IEEE(Year).
14. Laeng, B., S. Sirois, and G.J.P.o.p.s. Gredebäck, *Pupillometry: A window to the preconscious?* **7**(1): p. 18-27(2012).
15. Huang, C.-L. and C.-J.J.E.S.w.a. Wang, *A GA-based feature selection and parameters optimizationfor support vector machines.* **31**(2): p. 231-240(2006).
16. Bustos, R. and T. Gedeon. *Decrypting neural network data: a GIS case study*. in *Artificial Neural Nets and Genetic Algorithms* of Conference.: Springer(Year).
17. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* (2014).
18. Schaefer, A.M., S. Udluft, and H.-G.J.N. Zimmermann, *Learning long-term dependencies with recurrent neural networks.* **71**(13-15): p. 2481-2488(2008).
19. Hochreiter, S. and J. Schmidhuber. *LSTM can solve hard long time lag problems*. in *Advances in neural information processing systems* of Conference. (Year).
20. Gers, F.A., J. Schmidhuber, and F. Cummins, *Learning to forget: Continual prediction with LSTM.* (1999).
21. Hawkins, D.M.J.J.o.c.i. and c. sciences, *The problem of overfitting.* **44**(1): p. 1-12(2004).
22. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* **15**(1): p. 1929-1958(2014).
23. Sarle, W.S.J.C.s. and statistics, *Stopped training and other remedies for overfitting.* p. 352-360(1996).
24. CHE, M., X.-h. SUN, and Q.-q.J.M. HAN, *Disadvantages of the genetic algorithm compared with biologic evolution and improvement [J].* (2): p. 16(2006).
25. James, S.L., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a*

systematic analysis for the Global Burden of Disease Study 2017. **392**(10159): p. 1789-1858(2018).