DOI: 10.54254/2755-2721/95/2024CH0053

Enhanced robustness in machine learning: Application of an adaptive robust loss function

Jiuyang Wan[1,5,*], Tianyi Wang[2], Wancheng Li[3], Ziyao Wang[4]

[1]School of Computing and Data Science, Xiamen University Malaysia, Selangor, Malaysia
[2]The Hotchkiss School, Lakeville, US
[3]School of Mathematics and Statistics, University of New South Wales, Sydney, Australia
[4]Sauder School of Business, University of British Columbia, Vancouver, Canada

[5]lwc20010110@gmail.com
*corresponding author

Abstract. In the field of machine learning, the selection of a loss function plays a pivotal role in determining the training dynamics and generalization capability of models. Traditional loss functions such as mean square error (MSE) and cross-entropy are often not equipped to handle noisy and anomalous data effectively, which can result in models that perform poorly in practical scenarios. To address these shortcomings, this study introduces a novel adaptive robust loss function, enhanced by a tunable parameter α, which allows for flexibility in adjusting the robustness of the function according to the nature of the data being processed. Our research demonstrates that this new loss function significantly improves the performance of linear regression and multilayer perceptron models, particularly in environments laden with noisy and anomalous data. By adapting the parameter α, the function can cater to varying levels of data irregularities, thus enhancing the model's accuracy and reliability across diverse and complex data environments. This adaptive mechanism not only offers a substantial theoretical contribution to the understanding of robust loss functions but also provides a practical tool for machine learning practitioners to develop models that are resilient to data imperfections. The implications of this research are profound, suggesting a shift towards more adaptive and robust approaches in machine learning model development.

Keywords: Machine learning, loss function, linear regression, multilayer perceptron.

1. Introduction

1.1. Research Background

Among the many parts of machine learning, the loss function lies at its core. It not only defines the goal of the optimization problem but also directly influences how well the models learn on the training data and generalize on new data [1]. The relative difference between the predicted and actual values is quantified by the loss function, which is to some degree dependent on its form and properties to set up a model's optimization-based difficulty level for the learning algorithm and final performance. Appropriate loss function selection or design is an extremely important step when it comes to realizing an efficient machine learning model.

DOI: 10.54254/2755-2721/95/2024CH0053

1.2. Research Motivation

Although classic loss functions like mean squared error or cross-entropy loss perform very well in most machine learning tasks, they have significant limitations when handling noisy and anomalous data. When experimenting, these loss functions were found to be sensitive to outliers. In practice, data is always filled with noise and irregularities, and this calls for the loss function to enhance its fault tolerance for better model adaptability and stability in complex environments [2]. Therefore, how to develop and apply a loss function that can handle abnormal data effectively, so as to improve the robustness of the model, has become a main direction of machine learning research.

1.3. Research objectives and contributions

This study mainly aims to investigate and validate a general and adaptive robust loss function that adjusts its fitting parameters dynamically according to different data distributions, especially in situations with noisy and anomalous data. It particularly elaborates on the mathematical construction and theoretical foundation of the loss function, and then applies it in different machine learning models such as linear regression and multilayer perceptron (MLP) to find its significant advantages in improving the ability of the model in handling abnormal data. The major contributions of the research lie in the following aspects: It proposes a new loss function, mathematically describes it, and is capable of providing smooth transitions between different robustness requirements, hence enabling wider application flexibility than traditional methods. The efficiency of the loss function in handling noise and outliers for models like linear regression and multilayer perceptron shall be proved on their basis by conducting some specific experiments. The loss function has already proved to be efficient on real data, significantly in those application scenarios where the quality of the data is homogeneous; that is, it improved both predictive stability and accuracy for a model.

2. Overview of Loss Functions in Machine Learning

Loss functions measure the error between predicted outputs and actual values of a machine learning algorithm (probabilistic). They are crucial in guiding the optimization process to improve model performance by minimizing the discrepancy between predicted and observed values (review). Loss functions are typically used in classification and regression, classification loss functions include cross-entropy, hinge, ramp, pinball, and exponential loss functions (review). These functions usually help to determine how well a model distinguishes between different classes of data. While regression loss functions include mean squared error (MSE), absolute loss (MAE), Huber, and log-cosh loss functions (review). In regression, loss functions quantify how well a model's predictions align with continuous target values [3].

Robustness in the context of loss functions refers to the ability of the loss function to handle noisy data, outliers, or model deviations without being significantly affected. A robust loss function minimizes the impact of such adverse conditions, leading to more reliable and stable model performance. Traditional loss functions like Mean Squared Error (MSE) can be very sensitive to outliers because the error is squared, giving more weight to larger errors.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

In contrast, robust loss functions like Huber loss mitigate the influence of outliers. The Huber loss function is quadratic for small errors and linear for large errors, providing a balance between MSE and Mean Absolute Error (MAE).

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } |a| \le \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \tag{2}$$

DOI: 10.54254/2755-2721/95/2024CH0053

However, there is no relatively accurate method for selecting δ. The general method is to use the exhaustive method for cross-validation. This trade-off is often encountered in machine learning. If the loss function is too robust, some information provided by the data may be missed. If it is not robust enough, the model will be affected by some outliers. The adaptive robust loss function provides a method that treats the degree of robustness as an optimization problem and leads the computer to select for us.

3. Methodology

3.1. The General and Adaptive Robust Loss Function

Barron proposed a loss function in [4] which can be seen as a generalization of variate loss functions. The parameter α allows the function to adjust its shape, providing a continuum between different loss behaviors. It is essential in controlling the robustness of the general loss function. The α parameter can be tuned to fit the characteristics of different datasets and achieve a balance between robustness to outliers and smoothness for optimization.

$$f(x,\alpha,c) = \frac{|\alpha-2|}{\alpha}\left(\left(\frac{(x/c)^2}{|\alpha-2|}+1\right)^{\alpha/2}-1\right) \tag{3}$$

$c > 0$ is a scale parameter that controls the size of the loss's quadratic bowl near $x = 0$. The parameter $c$defines the sensitivity of the loss function with respect to the size of the error. Small values for c sensitively tune the loss function with small variations, while large ones do not specifically tune the function relatively closer to the large errors, allowing outliers to be as such.

Equation (4) gives some existing loss functions that we can recognize.

$$\rho(x,\alpha,c) = \begin{cases} \frac{1}{2}\left(\frac{x}{c}\right)^2 & \text{if } \alpha \to 2 \\ \log\left(\frac{1}{2}\left(\frac{x}{c}\right)^2+1\right) & \text{if } \alpha \to 0 \\ 1-\exp\left(-\frac{1}{2}\left(\frac{x}{c}\right)^2\right) & \text{if } \alpha \to -\infty \\ \frac{|\alpha-2|}{\alpha}\left(\left(\frac{(x/c)^2}{|\alpha-2|}+1\right)^{\alpha/2}-1\right) & \text{otherwise} \end{cases} \tag{4}$$

When α approaches 2, the function will penalize errors quadratically, similar to the behavior of an L2 Loss function. When α is set to 1, the function will behave similarly to a Charbonnier Loss, by transitioning between L1 and L2 behaviors [5]. When α is set to 0, the function will behave like the Cauchy loss, which grows logarithmically. When α is set to -2, the function will be similar to the Geman-McClure loss, by assigning smaller weights to large residuals. When α approaches negative infinity, the function will use an exponential term to minimize the effect of large residuals, similar to the Welsch loss.

During training, α is treated as a learnable parameter and optimized alongside the model weights by minimizing the negative log-likelihood (NLL) of the model's distribution. This approach prevents the trivial solution of setting α too low, which would excessively discount outliers.

3.2. Application in Linear Regression

Adapting a robust loss function in the linear regression modelling technique makes that model more tolerant toward noise and outliers. The classical mean square error (MSE) loss function works quite effectively in those datasets that are low in noise; however, it can suffer to a great extent when datasets contain many outliers. In other words, adaptive robust loss does not bring about insensitivity to small errors. Through a and c, adaptive robust loss allows the model to reduce sensitivity toward enormous mistakes, without losing sensitivity to small errors. From this perspective, adaptive robust loss is applicable to a wide variety of datasets. Experiments involving the adaptive robust loss function showed

DOI: 10.54254/2755-2721/95/2024CH0053

that the prediction made for datasets with noise was much more accurate and stable than that obtained by the standard MSE loss function. That allows changing the value of a, which defines how much the model is sensitive to the outliers, and hence, elevates an optimal overall prediction performance.

3.3. Application to Multilayer Perceptron (MLP)

Another area where significant advantages can be seen is the use of adaptive robust loss functions in multilayer perceptron (MLP) models. For model over-tuning due to wrong labels, problems from overfitting caused by label noise can let the traditional loss functions like cross-entropy become a troublemaker. In this case, flexibility of form in the adaptive robust loss function helps to get away with the problem. An explanatory experimental setting alternates the parameter values of a and c, all of which are to identify the role that the loss function plays on the classification accuracy performance of the MLP model under the different settings. Experimental results show that proper choices of parameters greatly reduce the overfitting phenomenon of the model and lead to good recognition of real labels, especially when a dataset has much label noise in it.

4. Experiments

4.1. Experiments on Linear Regression

We have implemented two models, a control model and an experimental model. For the control model, we have implemented linear regression with the classic MSE loss function. For the experimental model, we utilized the general robust loss function. The datasets used for testing were generated with numpy, specifically to set them at different outlier percentages.

By doing so, we hope to test both model's effectiveness with different outlier levels and examine whether the general robust loss function would be more effective. The outlier percentages were specifically set at 0%, 5%, and 20%.

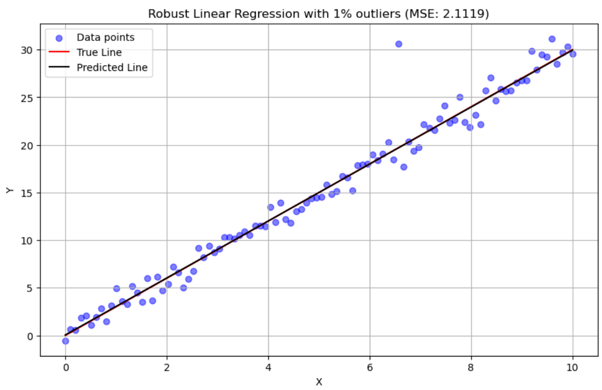

Figure 1. Robust model at 1% outlier level

As expected, with low outlier percentages, both the non-robust and robust models perform similarly well, as shown in Figures 1 and 2.

At a 5% outlier percentage, the inaccuracy of the non-robust model becomes apparent, as illustrated in Figures 3 and 4. The robust model maintains better accuracy, highlighting its effectiveness under these conditions.

When the proportion of outliers reaches 20%, the non-robust model's predictions deviate significantly, while the robust model shows a remarkable resistance to outliers, as depicted in Figures 5 and 6. These results underscore the robust model's ability to ignore outliers during the training process, thereby not deviating far from the true parameters. It is worth noting that the robust model has a larger MSE, because

Figure 2. Non-robust model at 1% outlier level

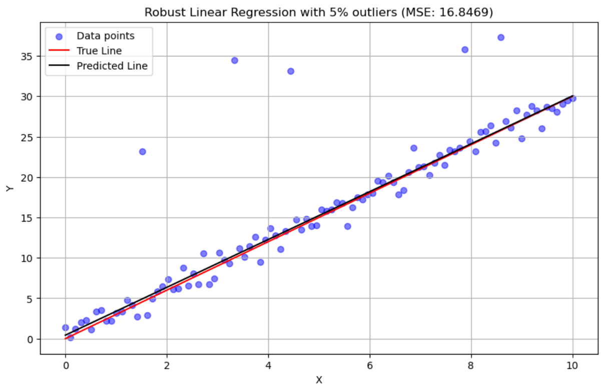

Figure 3. Robust model at 5% outlier level

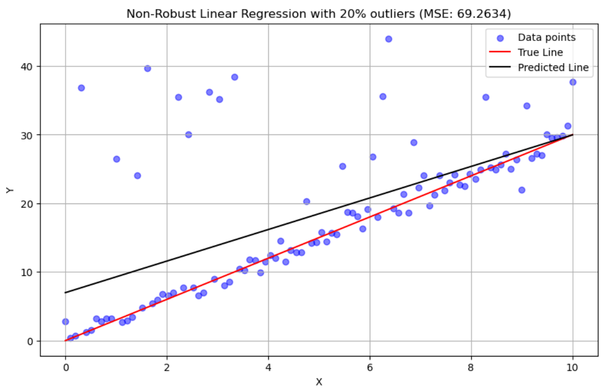

Figure 4. Non-robust model at 5% outlier level

in the ordinary linear regression training process, the optimization goal is to minimize the MSE. The robust model ignores the outliers during the training process, so it does not deviate too far from the true parameter.

Figures 7 and 8 further explore the impact of outliers and the optimization of the α parameter on model performance. The experimental results show that model accuracy significantly deteriorates with an increasing proportion of outliers, but optimal tuning of α can substantially enhance model robustness

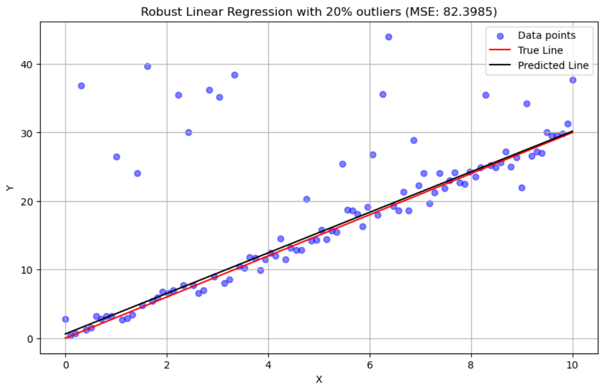

Figure 5. Robust model at 20% outlier level

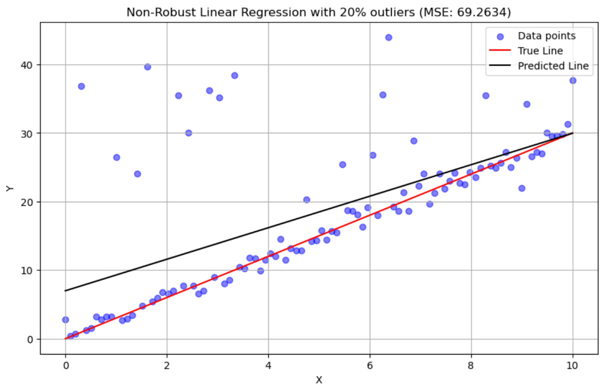

Figure 6. Non-robust model at 20% outlier level

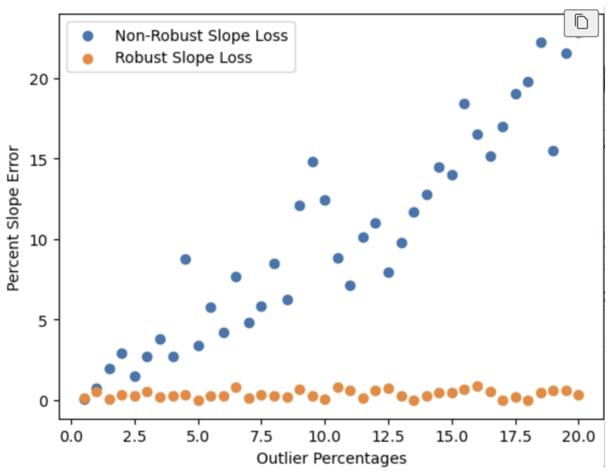

Figure 7. Comparison of percentage of error between two models

DOI: 10.54254/2755-2721/95/2024CH0053

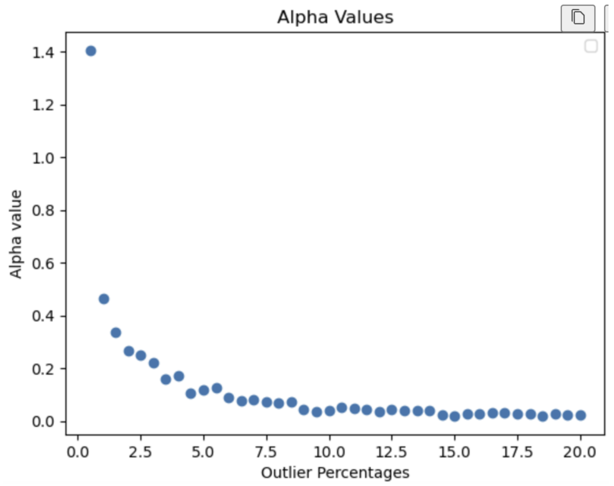

Figure 8. Adaptive value of α with respect to percentage of outlier

and accuracy in challenging data environments. Recall that, the value of α is adaptively selected during the training process. The smaller the alpha, the more robust the loss function. The α value in the figure also verifies this point of view. As the proportion of outliers increases, we need the loss function to become more robust.

4.2. Experiments on Multilayer Perceptron

In the multilayer perceptron experiments, an adaptive robust loss function is used, and a set of experiments are run to find out how different parameters affect model performance. The multilayer perceptron model applied in this experiment had 6 neurons in its input layer, corresponding to the number of features in the input data. The first hidden layer contains 4 neurons, the second 4, and there are a total of 3 neurons in the output layer, corresponding to the number of categories for the classification task. In this way, the structure is complex enough to catch nonlinear relationships but not too complex to avoid overfitting. More specifically, the layers are fully connected with one another, while the activation function utilized for ensuring nonlinear mapping ability is the ReLU.

Table 1. Performance metrics for each fold and their averages.

Fold	Sensitivity	Specificity	Precision	Accuracy	F-score
1	0.772	0.959	0.821	0.937	0.784
2	0.845	0.974	0.877	0.958	0.848
3	0.785	0.960	0.776	0.935	0.778
4	0.824	0.972	0.828	0.954	0.822
5	0.778	0.967	0.795	0.946	0.781
Avg	0.801	0.967	0.819	0.946	0.803

The table 1 summarizes the performance metrics of a Multi-Layer Perceptron classifier on a classification problem without outliers across five folds. The dataset is still synthetic. Sensitivity, Specificity, Precision, Accuracy, and F-score are reported for each fold, as well as their averages. The MLP demonstrates consistent performance with high Specificity (0.960 to 0.974) across all folds, indicating its effectiveness in correctly identifying negative cases. The Sensitivity ranges from 0.772 to 0.845, showing a slight variation in identifying true positives. Precision values (0.776 to 0.877) reflect the classifier's ability to correctly label positive predictions. Accuracy is consistently high, ranging from 0.935 to 0.958,

DOI: 10.54254/2755-2721/95/2024CH0053

with an average of 0.946, highlighting the model's overall reliability. F-scores, balancing precision and recall, vary between 0.778 and 0.848, with an average of 0.803, indicating robust performance. These metrics collectively suggest that the MLP is a reliable model for this classification task, maintaining strong performance across multiple evaluation criteria.

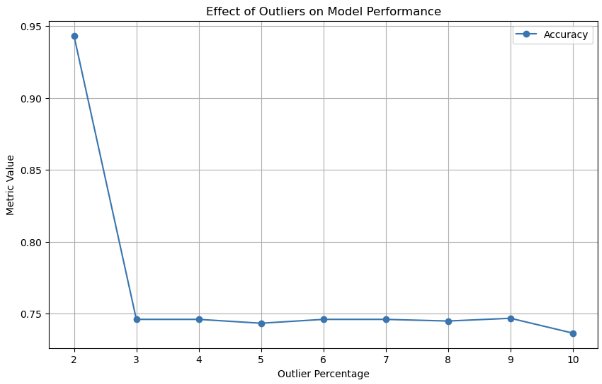

Figure 9. Accuracy of the model for different percentage of outlier

The above graph shows the experimental results on data containing outliers. Figure 9 illustrates how the introduction of outliers can drastically alter model performance, showing significant deterioration in accuracy with increasing proportions of outliers.It is observed from the graphs of experiments that model accuracy deteriorates significantly with an increasing proportion of outliers. At 2% of outliers, the accuracy is as high as 95%, while at 3% of outliers, the accuracy plummets to about 75%. This result underscores the high sensitivity of the model toward outliers under a standard MSE loss function and its failure to deal with a higher ratio of anomalous data.

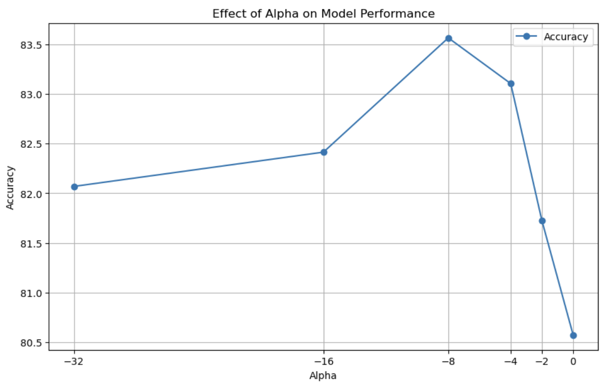

Figure 10. Accuracy of the model under different α

Figure 10 demonstrates the impact of the shape parameter α on model performance, with detailed experiments showing that adjusting α can optimize model accuracy, particularly in handling outliers effectively. Particularly, in the experimental setting, this included the 3% outlier dataset using an adaptive robust loss function, with α optimized over a wide range of values in order to quantify its individual impact on model performance. The results used accuracy as the key performance metric, which measured model efficacy under different settings of α. It can be seen that, while reducing α from 0 to -8, model accuracy dramatically goes up to about 83.5%, showing the best performance, thus indicating that setting α to -8 is an optimal way to resist outliers while remaining sensitive to normal data. However, from $\alpha = -32$, the accuracy of the model declines relatively flatly to about 82%, showing that over-adjustment to α will reduce the fitting ability to normal data, which means that the over-robustness of the loss function could affect model performance. The result of the experiment clearly shows the role of α in an adaptive robust loss function, particularly when treating complex datasets with outliers. An appropriate choice of α may greatly enhance the accuracy and robustness of the model in difficult data environments. However, the results also indicate that excessive tuning of α, which is too small, may turn out to be counterproductive and degrade model performance.

4.3. Comparative Analysis and Discussion

In this study, we investigate how far traditional loss functions generalize to generalized robust loss functions on handling different outlier ratios through comprehensive experiments with multilayer perceptron and linear regression models. Experimental results indicate that the robust loss function can greatly enhance the stability and accuracy of a model in outlier data.

In the experiments of linear regression, we consider a controlled model using the classical mean square error loss function versus an experimental model using the generalized robust loss function. The experimental setting includes various different proportions of outliers, from 0% to more than 15%. While the proportion of outliers increases, especially with outliers above 10%, the performance of the controlled model decreases significantly, and its predictive accuracy drops dramatically. In contrast, this experimental model retains high performance even at high outlier ratios, showing the generalization effectiveness of the robust loss function in high outlier ratio data.

Similarly, in the MLP model experiments, tuning of the shape parameter α of the robust loss function also shows its important role in controlling the stability and accuracy of the model. It shows the best performance on datasets with outliers when α is optimized to some specific value (-8 in this experiment), while too low or too high α negatively affects the performance of the model.

Such experiments prove that the design of robust loss function forms a critical basis for improving the robustness in the face of outlier data [6]. In the linear regression model, the generalized robust loss function helps hold the regression fitting line and predictive accuracy by effectively reducing the effect of outliers. Under suitable adjustment of the α value in the MLP model, it can improve the model's tolerance to noise and outliers to enhance the model's performance in complex data environments.

Combining the results of the experiment with linear regression and MLP, the generalized robust loss function performs better than the traditional MSE loss function for various outlier ratios, especially when facing highly outlier-rich datasets. These results show how important it is to consider the characteristics of the data and the presence of outliers in the design process of a machine learning model and the choice of a suitable loss function. The generalized robust loss function shows its effectiveness and capability of handling complex real-world noisy data by high resistance to outliers.

5. Conclusion

5.1. Summary Research

This study evaluates the application of generalized against adaptive robust loss functions within machine-learning models, particularly linear regression and multilayer perceptron models, using rigorous experiments. Evidently, from the results, it is stated that compared to the traditional squared mean error loss function, the robust loss function has a huge improvement in terms of model accuracy and stability when facing data sets with different proportions of outliers. This way, proper modulation of shape parameter α and scale parameter c in the loss function could effectively improve the robustness of the model to outlier data.

5.2. Research Significance and Application

This research is toward improving the loss function adaptability to the characteristics of data to optimize the performance of machine learning models with regard to noise and outliers. Applications of such an adaptive robust loss function allow for new solutions to problems within machine learning against the backdrop of non-standard and non-ideal datasets. Apply traditional loss functions, especially in financial, medical, or social media data analysis, where a large proportion of outliers may be expected due to human error, measurement error, or even fraud; expected results in prediction may not be achieved. Hence, the results of this study are of very practical significance when developing more accurate and reliable prediction models.

5.3. Future Research Directions

According to the result of the current study, several future research directions can be further expanded as follows: Firstly, research can be conducted on parameter optimization. The way how to automate the optimum selection of α and c can be further researched, perhaps leaving the machine learning algorithm itself to dynamically adjust such parameters to adapt to any changes in the data distribution during training. Secondly, it can be studied in terms of practical use cases. How well do general versus adaptive robust loss functions work on other types of machine learning models; for instance, more complex models such as Convolutional Neural Networks and Recurrent Neural Networks? Thirdly, cross-model applications can be attempted. Deploy these models in specific application scenarios, like sensing systems of self-driving vehicles or online fraud detection systems, to further estimate the performance of the models and optimize the model structure. The last is that deeper theoretical research can be conducted. We can aim at the position and roles these loss functions play in statistical learning theory and how they affect the generalization ability of models and learning theory. These studies will allow us not only to further the robustness of current machine learning models but also to give back to the improvement of theory and applications of machine learning.

Acknowledgment

Wancheng Li, Tianyi Wang, Jiuyang Wan and Ziyao Wang contributed equally to this work and should be considered co-first authors.

References

[1] Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of eeg-based brain–computer interface paradigms. *Journal of neural engineering*, 16(1):011001, 2019.

[2] A. Ghosh, H. Kumar, and P. S. Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[3] J. Qiu, Q. Wu, G. Ding, Y. Wang, and Y. Xu. A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1):67, 2016.

[4] Jonathan T Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4331–4339, 2019.

[5] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *ICIP (2)*, pages 168–172. IEEE Computer Society, 1994.

DOI: 10.54254/2755-2721/95/2024CH0053

[6] Z. Feng and W. Lu. Robustness analysis and improvement of deep learning for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3803–3811, 2018.