Comparative Study of Loss Functions in Personal Identification for Smartwatch Data

Examining Accuracy of Loss Functions in Personal Identification using Outlier Detection with Auto-encoders on Data from Smartwatches

Ege Yümlü[1]

Supervisor(s): David Tax[1]**, Arman Naseri Jahfari**[2] **Ramin Ghorbani**[3]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Ege Yümlü
Final project course: CSE3000 Research Project
Thesis committee: David Tax, Arman Naseri Jahfari, Ramin Ghorbani, Guohao Lan

Abstract

Smartwatches are equipped with sensors that allow continuous monitoring of physiological and physical activities, making them ideal sources of data for data analysis. However, accurately identifying individuals based on smartwatch data can be challenging due to the presence of outliers. Hence, outlier detection techniques play a crucial part in this context by identifying and handling these data points. Auto-encoders are one of the prominent ways to address outlier detection. Auto-encoders minimize a loss function to identify outlier samples. To explore the most optimal loss function for smartwatch data, this paper conducts a comparative analysis between three unsupervised loss functions, fused directional loss, mean square error, and regularized loss extracted from the current literature. The performance of three functions in personal identification is employed as the performance criteria due to the lack of outlier labels. The results indicate that the auto-encoder's performance in personal identification is slightly better than random guessing. The model struggled to effectively capture individual characteristics of the training data. This led to the outlier samples and non-outlier samples not being separable in the evaluation set. Consequently, the variation in the performance across and within a loss function was primarily influenced by the characteristics of the data rather than the model itself. Thus, the auto-encoder has limitations in personal identification, which led to an inconclusive comparison of the loss functions.

1 Introduction

Smartwatches are equipped with sensors that allow continuous monitoring of physiological and physical activities, making them ideal sources of data for data analysis [1]. However, accurately identifying individuals based on smartwatch data can be challenging due to the presence of outliers [2]. Hence, outlier detection techniques play a crucial part in this context by identifying and handling these data points.

Auto-encoders are a type of neural network that can be used for unsupervised learning. Auto-encoders compress data to a lower dimensional representation, to later decode it. The difference resulting from this process, between the input and the output, is the error identified for that point. The identified error, with a certain threshold, is used to identify outliers. The role of auto-encoders in anomaly detection by using the identified error has been explored in the current literature. Covid-19 detection using heart rate [3], identifying surface defects in the manufacturing process [4], and anomaly detection for tools under noises to name a few [5].

The paper on Covid-19 detection explores the use of contrastive loss, which involves splitting the data into two classes, with the help of another model, and scaling them accordingly. Another model is to classify the samples and outliers, making labeling necessary for the training data. In the second paper, which addresses surface defect inspection, regularization is leveraged in the loss function to prevent over-fitting, leading to clustered feature maps around a point. The third paper discusses anomaly inspection for tools under noises. The loss function, fused directional distance (FDD), combines cosine distance and euclidean distance to also take into both the distances between points and the respective angle between them. The implementation of auto-encoders in the aforementioned papers varies, particularly in the choice of loss function used. This raises the necessity for additional research to determine the most precise loss function for outlier detection using heart rate data.

The main research question is identified as "What is the most accurate loss function for outlier detection in personal identification using auto-encoders for smartwatch data?". Furthermore, the process of personal identification and its accuracy will be employed to validate the effectiveness of the outlier detection approach.

This raises the following further sub-questions: How can we determine the most accurate loss function for outlier detection in personal identification using auto-encoders on smartwatch data? What are the strengths and weaknesses of different loss functions, and how do they affect the performance of the auto-encoder model? The data used for the experiment is non-labeled, indicating that contrastive loss will not be the focus. Instead, the focus will be on regularized and FDD loss as potential solutions when answering the main and sub-questions.

The main contribution of this research is a comparative study of different loss functions for outlier detection in personal identification using auto-encoders on smartwatch data. The study evaluates the performance of different loss functions and provides insights into their strengths and weaknesses. The results of the study can inform the development of more accurate and reliable methods for outlier detection in personal identification based on smartwatch data.

The paper will explore first the methodology to give an overview to the reader on the topic, later followed by the responsible research section to show the ethical implications. The experimental setup employed with the results gathered will be the next chapter. Following the results, the discussion chapter will go over the gathered results and give an explanation. Finally, a conclusion will convey the main points to the reader.

2 Research Methodology

This chapter explores the research methodology adopted in identifying the most accurate loss function for outlier detection in personal identification using auto-encoders for smartwatch data. The methodology encompasses crucial components data collection and preprocessing, auto-encoder, loss functions, hyper-parameter optimization, and evaluation of their performance.

2.1 Data Collection and Preprocessing

Data used for training and evaluation of the model were sourced from the clinical trials called Machine Learning Enabled Time Series Analysis in Medicine or in short ME-TIME. The data is gathered from Fitbits. The data structure contains the following component: heart rate, step counter, time, and other omitted metadata. Since the study explores the use of heart rate, all other data was redacted, and not used.

Data preprocessing had three main steps:

- The original sample rate was variable, 0.2 Hz was identified as the most prevalent. Hence, the heart rate was re-sampled to 0.2 Hz, corresponding to once every 5 seconds.
- For the data points where the heart rate is missing from a maximum of 12 samples, the missing heart rate values have been linearly interpolated.
- For the data point where the heart rate is constant for more than 12 samples, the constant sequence is removed to provide more insights.

The data per individual was then split into window sizes determined in 4.1 with no stride.

2.2 Auto-encoder

A simple auto-encoder for unsupervised learning consists of two critical components: the encoder, and the decoder. The encoder takes a feature vector x as input, for the dimensionality to be reduced until the latent layer. The abstract, lower dimensional representation at the latent layer Figure 1 is then fed to the decoder, where it gets decompressed back to a reconstructed feature vector $\bar{x}$ with the original input dimensionality. The encoder has a series of linear layers and activation functions, to be able to represent the data distribution of the training samples in the latent layer. The auto-encoder is symmetric, leading to the decoder mirroring the encoder layers in reverse order. The general structure of an auto-encoder is illustrated below to give a better insight.

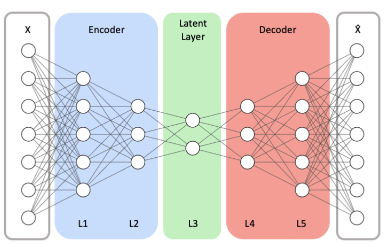

Figure 1: An example auto-encoder model architecture with symmetrical encoder and decoder networks

The auto-encoder is trained with a loss function to optimize the reconstruction error, for the given training data. This way samples from the original distribution are represented with a smaller reconstruction error, compared to the non-trained samples which makes them easier to identify.

2.3 Loss Functions

An auto-encoder is typically optimized by minimizing the reconstruction error. The main loss functions that will be explored are mean square error (MSE), fused directional distance loss (FDD), and regularized loss.

Mean Square Loss

The most basic approach to the loss function is the MSE. The loss is depicted as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1}$$

where

- n is the number of samples
- $\bar{x}$ is the output for the input vector x
- x_i is the sample at the index i

Due to its simplicity in computation and interpretation, the mean square error is frequently employed in a wide range of applications. Its sensitivity to outliers, however, is a significant disadvantage since the model may become unduly preoccupied with fitting the outliers at the expense of the majority of the data. Hence euclidean distance is difficult to accurately describe the similarity of complex feature spaces [6]. As a result, the MSE loss function will be used as a baseline for comparison in this study with other loss functions.

Fused Directional Distance Loss

In order to effectively suppress the effect of noises and increase the model's robustness for anomaly identification, the FDD loss function combines mean squared error (MSE) and cosine similarity (CS), which can quantify distribution similarity between data from the perspectives of angle and distance. The FDD loss function is used as described in the following paper [5]. The formula of cosine similarity, and FDD is below respectively:

$$CS = \frac{\sum_{i=1}^{n}\bar{x}_i \cdot x_i}{\sqrt{\sum_{i=1}^{n}(\bar{x}_i)^2} \times \sqrt{\sum_{i=1}^{n}(x_i)^2}} \tag{2}$$

$$FDD = MSE + \lambda\frac{(1 - CS)}{2} \tag{3}$$

where

- n is the number of samples
- x_i is the sample at the index i
- $\bar{x}_i$ is the predicted value for the input x_i
- λ is an adjustable hyper-parameter, which is used to balance the weight of distance and angle difference between data.

Regularized Loss

Regularized Loss involves adding a penalty to the term on the loss function to prevent over-fitting. Regularization can be added to the initial loss function in the context of auto-encoder to produce feature vectors that are densely packed during training. A traditional auto-encoder calculates the reconstruction error through back-propagation, resulting in a wide range of feature maps. The regularized loss promotes the tendency of feature maps to cluster around a central point [5]. Clustering along a nucleus produces more comparable feature vectors for comparable input data. By addressing the problem of feature maps with a large range in conventional auto-encoder, this method enables more accurate representations of the input data [5]. The formula is as follows from the respective source applying tight regularization:

$$L^l(X) = \frac{1-\lambda}{N}\sum_{i=1}^{N}|x_i - \bar{x}_i^{(l)}|^2 + \lambda\frac{1}{N}\sum_{i=1}^{N}(\max\{d_i(l) - \overline{d}^{(l-1)}, 0\})^2 \tag{4}$$

where

- l is the current epoch
- $\bar{x}_i$ is the predicted value for the input x_i
- $\overline{V}^{(0)} = 0$
- $d_i(l) = ||V_i^{(l)} - \overline{V}_i^{(l-1)}||$, where $\overline{V}_i^{(l-1)}$ is the mean feature vector at epoch $l-1$
- $\overline{d}^{(l-1)} = \frac{1}{N}\sum_{i=1}^{N} d_i^{(l-1)}$
- λ is the penalty factor, and λ is in the range [0,1]

The loss function behaves as a conventional auto-encoder with MSE when $\lambda = 0$. With the increase in the value of $\lambda > 0$, the second part of the loss function makes the feature vector of each sample to be as close to the center $\overline{V}$

2.4 Hyper-parameter Optimization

Hyper-parameter optimization is an important step to ensure valid comparisons and achieve optimal performance for the given model. In this study, hyper-parameter optimization was conducted using the tool Weights and Biases with the sweep functionality [7]. Sweep is a functionality that allows the exploration of various values and their respective performances. To perform the exploration, a random uniform distribution was assumed for the values in a given range. With the values selected, a grid search is performed. This approach ensured a comprehensive search of the space, allowing a thorough analysis of the configurations.

2.5 Evaluation Methodology

The performance evaluation of the auto-encoder model will be conducted on a dedicated evaluation set, consisting of an equal split between outlier and normal samples. The normal samples are obtained from the individual on whom the model was trained, while the outlier samples are obtained from individuals who were not part of the training data. Each optimized auto-encoder for each loss will be evaluated on its ability to detect outliers. The evaluation metric used for this task is accuracy and f1 score.

To mitigate the effect of random weight initialization, the experiments averaged independent runs. An average of 10 runs was deemed to be adequate, as there was a minimal change compared to using 20.

The performance of the four different models in outlier detection will be evaluated using a comparative analysis with the accuracy and f1 score. Additionally, visualization tools will be used to provide insightful representations to provide a more comprehensive understanding of the models' performance and effectiveness in outlier detection.

3 Responsible Research

This chapter includes the responsible research considerations and their subsequent consequences for the given study. First, ethical considerations are presented in their respective chapter. Following the ethical considerations, the reproducibility of the experiment is evaluated.

3.1 Ethical Considerations

The use of personal data from smartwatches for research raises imported ethical considerations. In this study, the privacy concerns and the risks associated with collecting and analyzing are acknowledged by the researcher. To address these concerns, several measures were taken to ensure participants' privacy and anonymity. Firstly, the data was obtained through clinical trials of Machine Learning Enabled Time Series Analysis in Medicine (ME-TIME) with proper ethical approvals and consent from participants. Additionally, we ensured data anonymization by removing any personally identifiable metadata and non-used information from the data set.

The potential bias in the data collection process is recognized, as the data may not represent a diverse population. While we made efforts to mitigate bias by including a variety of participants, it is important to acknowledge this limitation in interpreting the findings. Future research should aim to include more diverse data sets to enhance the generalizability of the results.

3.2 Reproducibility

The repeatability of the experiment relies on five main points: data preprocessing, auto-encoder architecture, loss functions, hyperparameter optimization, and evaluation. In section 2.1, the data preprocessing methods are identified and discussed in detail, including the source of the data. Likewise, comprehensive information on the auto-encoder architecture such as the number of layers, activation functions, hidden units, and the specific parameter values. The loss function formulas have been shared, along with the respective source papers, to ensure transparency and reproducibility. The hyper-parameter optimization methods and values are shared through informative graphs and tables. Finally, evaluation methodology has been explored, detailing the metrics and the techniques used to evaluate the performance of the models. The link for the code has also been provided in appendix B.

By providing all these details, transparency and reproducibility are adhered to. One point to mention would be that the individuals' respective ids are redacted in the graph, and converted to their respective order. Reproducibility does not take prominence over an NDA, resulting in the given redacted data.

4 Experimental Setup and Results

This chapter presents the experimental setup for implementing the auto-encoder and provides detailed explanations for the choices made. The optimization process and the resulting outcomes are then presented for discussion.

4.1 Auto-encoder Optimization

The Autoencoder build process had many layer structures to explore to find the optimal configuration. Several different layers were tested to determine the best-performing model based on the reconstruction error, with the baseline loss function MSE. The window size was determined to be 100, to include almost 10 minutes of heart rate data and heart rate variability.

The initial structure consisted of a linear layer, responsible for converting the window size to a lower-dimensional representation. The resulting reconstruction error per perceptron amount was plotted to visualize the relationship between the two.

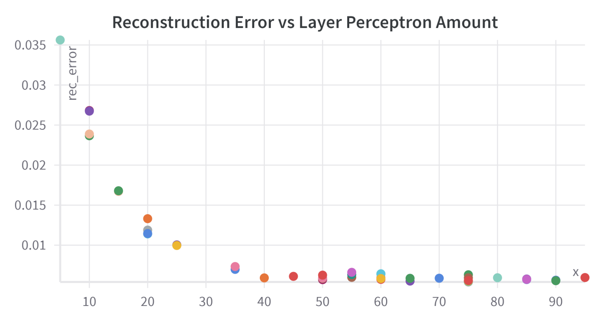

Figure 2: Reconstruction Error plotted against perceptron amount in layer one as variable x for the auto-encoder with a mean square error

The reconstruction error stabilized after reaching the perceptron amount of 40. However, to ensure greater stability across all loss functions, perceptron amount 55 was employed. This choice was based on multiple trial runs, which consistently resulted in stable results.

Following the initial layer, a second layer was introduced. The same techniques explored in the first layer were applied to the second layer. The results of the visualization can be seen below.

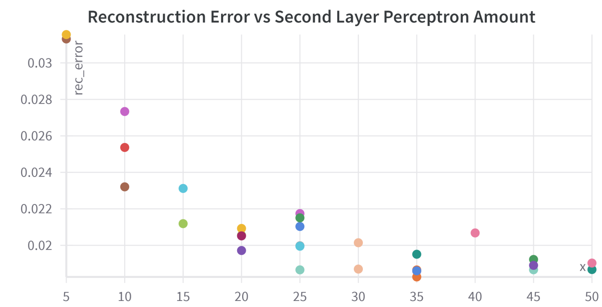

Figure 3: Reconstruction Error plotted against perceptron amount in layer two as variable x for the auto-encoder with a mean square error.

The total reconstruction loss for two layers ended up being higher compared to only using one layer. However, the objective of this study is personal identification rather than solely minimizing the reconstruction error. The inclusion of a second layer makes the auto-encoder more specialized for the training data, which would lead to a higher ability to do personal identification. As a result, the second layer perceptron amount of 35, is identified as the optimal value from the graph. The final structure is two linear layers of 100-55 and 55-35 were determined as the final structure to use throughout the paper.

4.2 Hyper-parameter Optimization

MSE

The hyper-parameter optimization was done to determine the learning rate. Using the Weights and Biases tool, a uniform distribution of learning rates between 0.01 and 0.00001 is explored. The Weights and Biases sweep resulted in the following graph of learning rate and reconstruction errors.

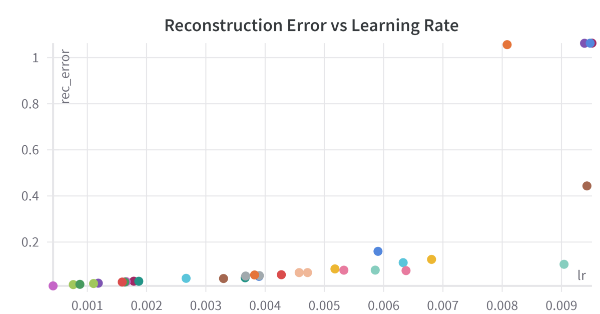

Figure 4: Reconstruction Error plotted against learning rate with MSE loss function and random search.

Following the random search, a grid search was conducted with the values 0.0005 to 0.004 incremented by 0.0005 as these seemed to yield the lowest reconstruction error. The results from the grid search can be seen in the figure below.

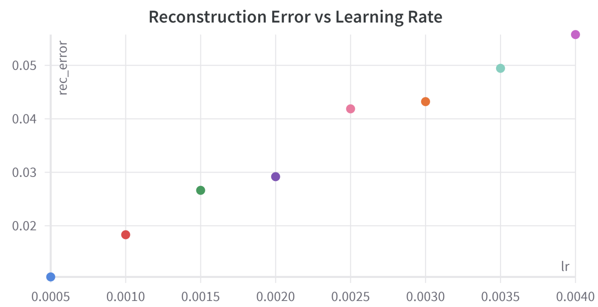

Figure 5: Reconstruction Error plotted against learning rate with MSE loss function and grid search.

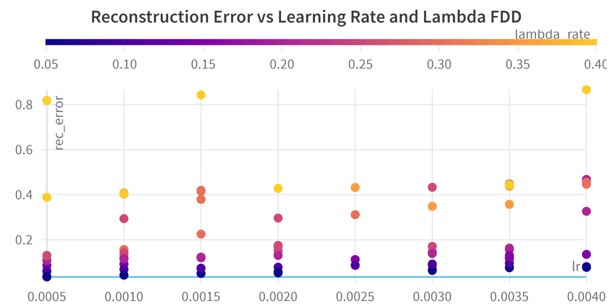

Figure 7: Reconstruction Error plotted against learning rate and lambda with FDD loss function and griid search.

For the sweep, the minimum reconstruction error was 0.0104 for the learning rate value of 0.0005. For this learning rate, loss per epoch seemed to be as expected. As a result, This value will be used for all future evaluations including the auto-encoder with MSE loss.

FDD Loss

FDD Loss hyper-parameters were explored using the Weights and Biases tool. In combination with the learning rate, the lambda parameter for the loss function was optimized to minimize the reconstruction error. The sweep explored values from a uniform distribution between values 0.01 and 0.00001 for the learning rate and 0.0 and 1.0 for the λ respectively.

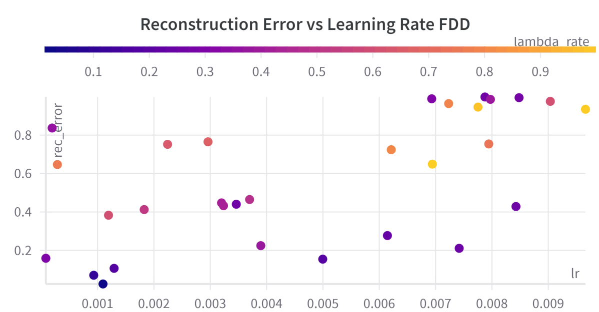

Figure 6: Reconstruction Error plotted against learning rate and lambda with FDD loss function and random search.

Following the random search, a grid search was conducted with the values 0.0005 to 0.004 incremented by 0.0005 for the learning rate, and the values 0.05 to 0.4 incremented by 0.05 for the λ as these seemed to yield the lowest reconstruction error. The results from the grid search can be seen in the figure below.

The optimal values identified were 0.0005 for the learning rate and 0.05 for the λ leading to a reconstruction error of 0.035. Training loss per epoch also behaved as usual for the selected hyper-parameters. As a result, these values will be the identified hyper-parameters for FDD loss It should be pointed out that this is a relatively high value compared to the rest of the loss functions reconstruction errors. This could be due to the nature of the FDD function or the data being harder to interpret for the FDD loss function.

Regularized Loss

Regularized Loss hyper-parameters to explore were lambda and learning rates similar to the FDD loss. As a result, the same process was applied to the regularized loss with the same parameters and their respective distributions.

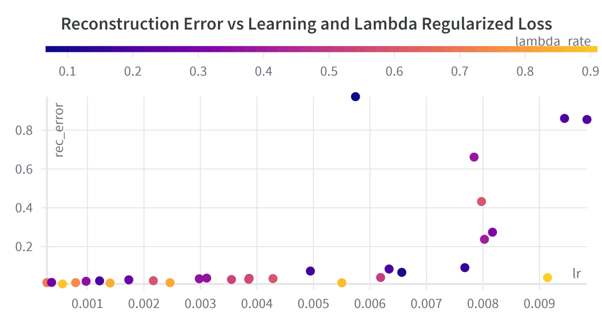

Figure 8: Reconstruction Error plotted against learning rate and lambda with regularized loss function and random search.

Following the random search, a grid search was conducted with the values 0.0005 to 0.004 incremented by 0.0005 for the learning rate, and the values 0.1 to 0.7 incremented by 0.1 and 0.7 to 0.9 incremented by 0.05 for the λ as these seemed to yield the lowest reconstruction error. The results from the grid search can be seen in the figure below.

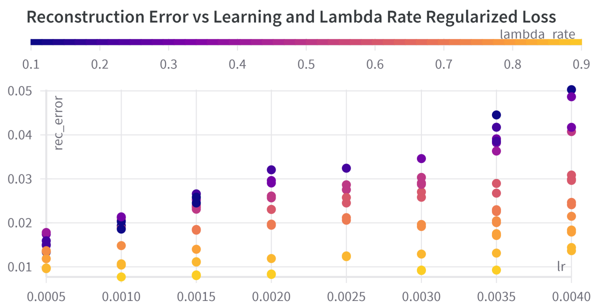

Figure 9: Reconstruction Error plotted against learning rate and lambda with regularized loss function and grid search.

The optimal values were 0.001 for the learning rate and 0.9 for the lambda. The training loss per epoch behaved as expected for the selected values. Hence, these values will be the identified hyper-parameters for regularized loss.

Threshold Optimization

The threshold for a given loss function is the final parameter to be optimized. For each Loss function, the percentile is iterated per training person to see the total accuracy of the given percentiles over the remaining 31 individuals. The evaluation set is iterated with different outliers to ensure the generalizability of the threshold. The table for all four thresholds identified can be seen below.

	RMSE	FDD	Regularized
1	5	5	90
2	40	45	35
3	90	75	70
4	10	90	90
5	35	50	35
6	90	5	5
7	15	15	10
8	25	25	25
9	40	70	60
10	90	85	85

Figure 10: Thresholds identified per individual across all loss functions for the highest personal identification accuracy

4.3 Evaluation Process

To evaluate the model, 10 random individuals were selected. One of the individuals is selected as the non-outlier sample from a selection of 32. The auto-encoder gets trained by the data of the non-outlier. After, all the individuals are iterated over to get the evaluation metrics accuracy, outlier accuracy, normal accuracy, and F1 score. The non-outlier is iterated to ensure every sample gets to be the non-outlier sample. The average accuracy per individual for all loss functions can be seen in the graph below.

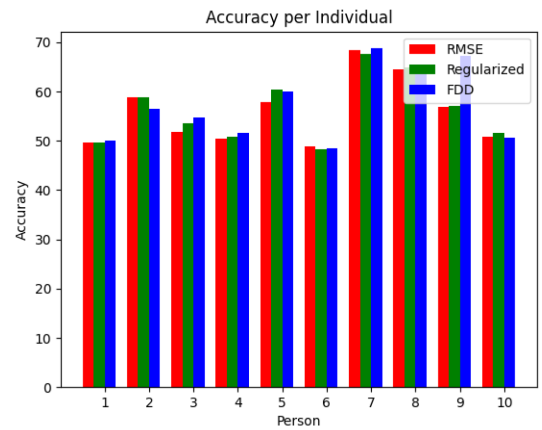

Figure 11: Accuracy identified per individual across all loss functions

5 Discussion

The experimental results yielded an accuracy of around 56% for personal identification for MSE loss. When the losses were alternated, the accuracy varied by a negligible 1-2% between the models. This suggests that the model is performing barely better than random selection by around 6%, which means that model built is an ineffective personal identifier.

To investigate these findings, a series of visualizations were performed, plotting the reconstruction error for the evaluation sets. Two graphs from individual 10 are presented below, one illustrating the reconstruction error before training and the other after training for the evaluation set. Upon analyzing these graphs, it becomes evident that the auto-encoder does not specialize for the given individual in the training set. It reduces the reconstruction error for both sets, indicating that there is no distinct threshold that can be implemented for optimal outlier detection. Consequently, the auto-encoder proves to be an unsuitable method for outlier detection in the context of personal identification.

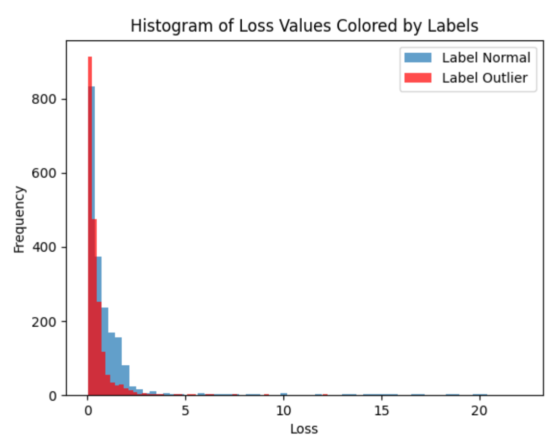

Figure 12: Evaluation Set Frequency Graph with No Training for Individual 10

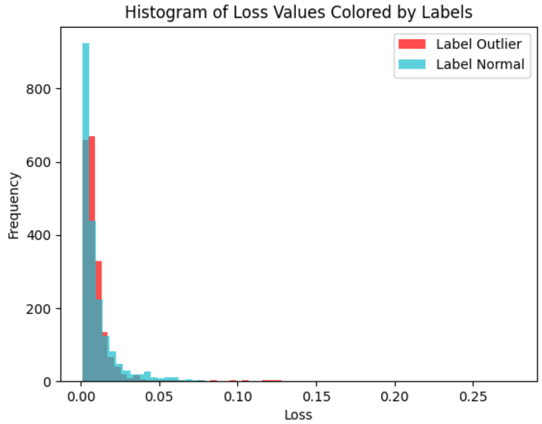

Figure 13: Evaluation Set Frequency Graph with Training for Individual 10

To investigate the causes behind the inadequate identification, a separate experiment was conducted. The hypothesis was that the similarity between the extracted data might have been due to inactive data. Therefore, a step filter was applied to both the training and evaluation data. The full results of this experiment, available in Figure 19, indicate that the overall average accuracy slightly decreased for the individuals by around 2-3%. However, certain individuals who performed poorly showed a small improvement. Notably, individual 111, which is the best performer in terms of accuracy for all loss functions, observed a drastic drop from 68% to 53%. After further exploration, as seen by the graph which compares the non-trained data for the individual 7 with and without heart rate, it becomes evident that the accuracy of the identification is affected more by the characteristics of the data rather than the underlying model itself.

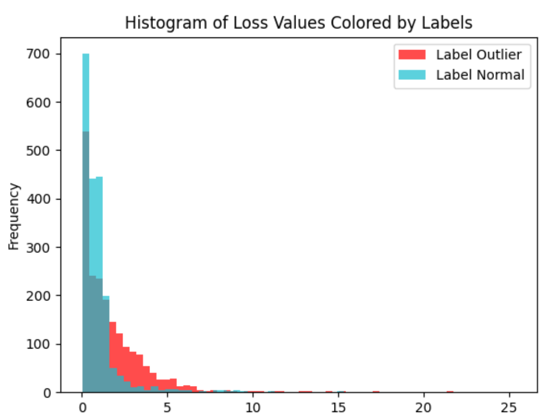

Figure 14: Evaluation Set Frequency Graph with No Training for Individual 7

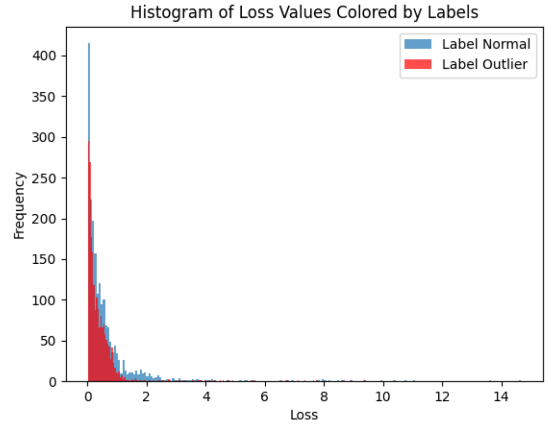

Figure 15: Evaluation Set Frequency Graph with No Training and Step Filter for Individual 7

6 Conclusions and Future Work

This chapter states the final conclusion drawn from the prior chapter. Later, the limitations of the current paper and recommendations for future papers are stated.

6.1 Conclusions

The research question of this study was focused on identifying the most accurate loss function for outlier detection using an auto-encoder. Since the data lacked explicit labels for outlier detection, the task of personal identification was chosen as the criterion. However, the auto-encoder struggled to effectively capture and leverage the individual characteristics present in the training data, leading to the non-separability of the classes in the evaluation set.

The performance of the auto-encoder, while slightly better than random guessing, did not result in significant accuracy in personal identification. Consequently, the variation across different loss functions was mainly influenced by the characteristics of the data, rather than the algorithm. This suggests that the auto-encoder exhibited limited ability in effectively identifying outliers for personal identification. The inability of the auto-encoder to distinguish individuals within the training data led to a lack of clear separation. As a result, the comparisons between the loss functions were deemed inconclusive. To conclude, the auto-encoder algorithm built, in its current form, results in low efficiency and effectiveness in outlier detection for personal identification.

6.2 Limitations and Future Work

Limitations include the simplicity of the auto-encoder architecture due to the time constraints of the study to provide a detailed comparison between the loss functions, future research should explore more complex models, with different layers to capture the intricacies of the data. The data source for the study was limited to ME-TIME data and Fitbits from a single clinical trial. This could limit the generalizability of the conclusions. In future research, the data set should be expanded with multiple sources to ensure a broader and more varied representation of data to enhance the validity of the given conclusion.

Furthermore, the data preprocessing techniques utilized, such as re-sampling the heart rate to a fixed rate or linearly interpolating the missing values have limitations. These methods might not fully capture the complexities present in the real-time data that could affect the accuracy of the model. Future investigations should explore the data preprocessing methods that can preserve the information and dynamics of the original data better.

In terms of future work, a key suggestion is to conduct research determining the optimal loss function for outlier detection using labeled data, which was not present in this study. By incorporating labeled data, it would be possible to evaluate the performance of outlier detection independent of personal identification. Additionally, a contrastive loss, which was not implemented due to a lack of labeled data and time constraints, should be explored and compared with the other loss functions. This could provide valuable insights into a comparison of its effectiveness within the outlier detection.

A Auto-encoder Experimental Results

The rows that have N/A did not have enough samples with the respective filter to match the batch size.

MSE Loss Summary Table

ID	Rec Error	Percentile	Accuracy	Normal Accuracy	Outlier Accuracy	F1 Score
1	0.0167	5	49.70%	89.45%	9.49%	63.90%
2	0.0162	40	58.92%	59.53%	58.30%	59.17%
3	0.0145	90	51.76%	86.60%	10.80%	62.80%
4	0.0112	10	50.50%	75.50%	93.45%	82.94%
5	0.0168	35	57.93%	51.53%	64.35%	55.06%
6	0.018	90	48.85%	96.00%	1.70%	65.24%
7	0.011	15	68.39%	50.92%	85.85%	61.70%
8	0.0212	25	64.58%	53.02%	76.15%	59.95%
9	0.013	40	56.85%	42.05%	71.65%	49.35%
10	0.01	90	50.79%	89.47%	12.10%	64.51%
	0.01486		55.83%	69.41%	48.38%	62.46%

Figure 16: Summary of Runs per Individual with MSE

FDD Loss Summary Table

ID	Rec Error	Percentile	Accuracy	Normal Accuracy	Outlier Accuracy	F1 Score
1	0.0417	5	49.97%	23.11%	95.70%	56.26%
2	0.0412	45	56.44%	60.39%	52.50%	56.44%
3	0.0409	75	54.77%	82.04%	27.50%	54.77%
4	0.0359	90	51.63%	92.85%	10.40%	51.63%
5	0.0421	50	60.03%	70.60%	49.45%	60.03%
6	0.0428	5	48.53%	1.57%	95.50%	48.53%
7	0.0366	15	68.78%	51.20%	86.35%	68.78%
8	0.0457	25	63.88%	52.15%	75.60%	63.88%
9	0.0392	70	67.34%	71.68%	43.00%	60.67%
10	0.0348	85	50.71%	84.42%	17.00%	50.71%
	0.04009		57.21%	59.00%	55.30%	57.17%

Figure 17: Summary of Runs per Individual with FDD

Regularized Loss Summary Table

ID	Rec Error	Percentile	Accuracy	Normal Accuracy	Outlier Accuracy	F1 Score
1	0.0186	90	49.76%	93.42%	6.10%	65.03%
2	0.018	35	58.95%	63.75%	54.15%	60.83%
3	0.0176	70	53.55%	83.02%	29.05%	65.38%
4	0.0127	90	50.87%	93.70%	8.05%	65.60%
5	0.0194	35	60.35%	62.91%	57.80%	61.34%
6	0.1986	5	48.40%	2.20%	94.60%	4.09%
7	0.01208	10	67.70%	47.36%	88.05%	59.46%
8	0.0224	25	64.89%	58.63%	71.15%	62.55%
9	0.0153	60	57.00%	67.50%	46.50%	61.09%
10	0.0109	85	51.67%	94.12%	8.90%	66.00%
	0.034558		56.31%	66.66%	46.44%	57.14%

Figure 18: Summary of Runs per Individual with Regularized Loss

MSE Loss Steps Filter Summary Table						
ID	Rec Error	Percentile	Accuracy	Normal Accuracy	Outlier Accuracy	F1 Score
1	0.0189	65	54.02%	62.94%	45.10%	57.79%
2	N/A	N/A	N/A	N/A	N/A	N/A
3	0.0138	25	54.46%	33.47%	75.45%	42.36%
4	N/A	N/A	N/A	N/A	N/A	N/A
5	N/A	N/A	N/A	N/A	N/A	N/A
6	N/A	N/A	N/A	N/A	N/A	N/A
7	0.0141	90	53.19%	89.65%	16.75%	65.70%
8	0.0212	25	64.58%	53.02%	76.15%	59.95%
9	0.0146	5	46.99%	3.93%	90.05%	6.90%
10	0.0174	85	53.59%	88.79%	18.40%	65.68%
	0.016667		54.47%	55.30%	53.65%	49.73%

Figure 19: Summary of Runs per Individual with MSE and Step Count Filter of 100

B Auto-encoder Experimental Results

The code for the auto-encoder used for this paper can be found in the following link: https://github.com/egeyumlu-cl/thesis

References

[1] CHEN Xiao-Yong, YANG Bo-Xiong, ZHAO Shuai, DING Jie, SUN Peng, and GAN Lin. Intelligent health management based on analysis of big data collected by wearable smart watch. *Cognitive Robotics*, 3:1–7, 2023.

[2] Seyed Bagher Estaghvirou, Joseph O. Ogutu, and Hans-Peter Piepho. Influence of outliers on accuracy estimation in genomic prediction in plant breeding. *G3 (Bethesda)*, 4(12):2317–2328, 10 2014.

[3] S. Liu, J. Han, E. L. Puyal, S. Kontaxis, S. Sun, P. Locatelli, J. Dineley, F. B. Pokorny, G. D. Costa, L. Leocani, A. I. Guerrero, C. Nos, A. Zabalza, P. S. Sørensen, M. Buron, M. Magyari, Y. Ranjan, Z. Rashid, P. Conde, C. Stewart, A. A. Folarin, R. J. Dobson, R. Bailón, S. Vairavan, N. Cummins, V. A. Narayan, M. Hotopf, G. Comi, B. Schuller, and R. A. D. A. R. . N. S. Consortium. Fitbeat: Covid-19 estimation based on wristband heart rate using a contrastive convolutional auto-encoder. *Pattern Recognition*, 123, 2022. Cited By :10.

[4] D. . Tsai and P. . Jen. Autoencoder-based anomaly detection for surface defect inspection. *Advanced Engineering Informatics*, 48, 2021. Cited By :35.

[5] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. *Robotics and Computer-Integrated Manufacturing*, 79, 2022. Cited By :34.

[6] A. Sethi, M. Singh, R. Singh, and et al. Residual codean autoencoder for facial attribute analysis. *Pattern Recognition Letters*, 119:157–165, 2019.

[7] Weights & Biases. W&B Sweeps Documentation, 2023.