Exploring Deep Learning for Cyclist Emotion Recognition

FLORIAN ANDREI MARIA, University of Twente, The Netherlands

This study explores the application of Recurrent Neural Networks (RNN) to recognise emotions using multimodal physiological signals from cyclists, specifically heart-rate variability (HRV) and electrodermal activity (EDA). These signals are used to develop prediction models(such as LSTM and GRU) that can be deployed on bike-mounted edge devices. This approach aims to enhance urban cycling safety by enabling timely adaptations. The study evaluates the feasibility of these models on edge devices and provides recommendations for their effective deployment to reduce accidents and improve cycling safety.

Additional Key Words and Phrases: "Emotion Recognition", "Deep Learning", "Physiological Data".

1 INTRODUCTION

1.1 Background

1.1.1 Cycling safety in the Netherlands. In the Netherlands, cycling is not just a leisure activity, it is the fundamental mode of transportation. This is evident from the fact that it is the number one country by bicycle use in Europe [4, 24], and has a strong infrastructure to support the millions of cyclists in the country. Despite that, there is evidence that cycling is becoming increasingly dangerous [9, 15, 23]. The number of cyclists seriously injured in traffic incidents has increased by over 30% in recent years [15], particularly the older cyclist being the number one category in mortal accidents [27], due to slower reaction time and increased vulnerabilities to serious accidents. An experiment in Utrecht found that many women feel unsafe riding alone at night [3]. This perception of danger can influence the levels of anxiety and stress and further compromise safety. A systematic Literature review of 10 years of cyclist safety research shows us a growth of the field that is anticipated to rise more. It also the necessity for data mining, more specifically mentioning it as "the last methodological frontier to explore cyclist crash data by searching for structures, commonalities, and hidden patterns or rules". The current studies also lack more research about emerging topics like e-bikes and are more focused on themes such as crash severity and safety equipment[19]. Based on the studies, this research will delve deeper into the potential that emotion recognition can have.

1.1.2 Emotion recognition from wearables. Recent studies have successfully employed wearable wristbands to monitor emotional responses [8, 20]. These devices, capable of measuring physiological data such as Heart-Rate Variability (HRV) and Electrodermal Activity (EDA), facilitate a possibility of enhancement for cyclist safety. While various libraries are available for analysing physiological data, pre-processing, and generating features [6, 11, 12] the practical application of these in cycling environments remains underexplored.

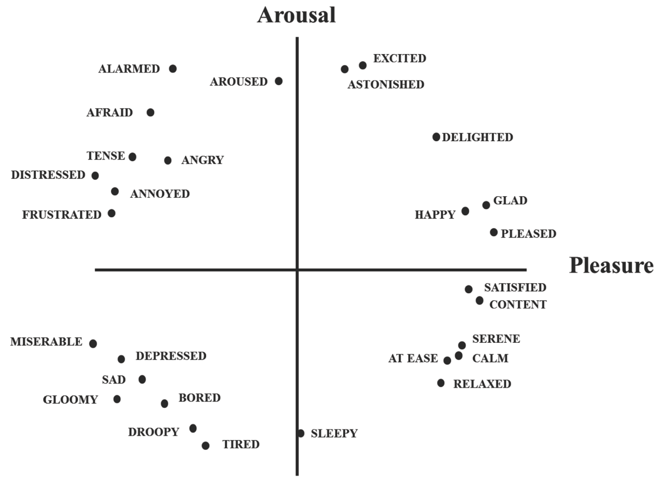

Fig. 1. Affective model

1.1.3 Affective model. Systems for recognising and interpreting emotions in humans utilise a framework called the affective model. In order to precisely identify and assess emotional states, this model often collects and analyses data from a variety of sources, including body language, speech tones, facial expressions, and physiological signals (such as the EDA, HRV, Electroencephalogram - EEG, and Electrocardiogram -ECG) [22]. The goal of affective computing, which includes these concepts, is to develop computers that can communicate with people naturally and efficiently way by being aware of their emotional states [22].

Monitoring the valence of cyclists can provide critical insights into their behaviour, potentially predicting and preventing accidents caused by stress, anxiety, or inattentiveness. As shown in the 1, the affective model frequently maps emotions on a two-dimensional plane determined by arousal, which is the physiological and psychological state of being awake or responsive to stimuli, and valence, which is the natural attractiveness or adverseness of an event. Emotions like joy, on the other hand, are characterised by both high arousal and high valence, whereas emotions like rage are characterised by low valence [7]. With a systematic approach to recognising emotions, this quadrant model assists in the classification and comprehension of various emotional states.

1.2 Problem Statement

The current approach to safety in the cycling area reflects this problem and is heavily focused on measures such as infrastructure and regulatory design enhancements [5]. While such measures prove useful, they are somewhat limited by nature in terms of their ability to respond to the immediate, individualised needs of cyclist in the ever-changing urban environments. The advent of consumer-friendly wearable technology has enabled the real time collection

TScIT 41, July 5, 2022, Enschede, The Netherlands

of physiological and emotional data, presenting a significant opportunity to enhance cyclist safety. For instance, heightened anxiety levels can lead to slower reaction times and increased risk-taking behaviours, critical factors in accident scenarios [13]. However, this potential remains largely unexplored in most safety frameworks.

The following research questions aim to guide this exploration and assess the feasibility and effectiveness of these innovative approaches:

1.3 Research Questions

Main Question: *How effectively can RNN models classify cyclists' valence states using multimodal physiological data for potential deployment on bike-mounted edge devices?*

Model Performance

(1) What are the accuracy and reliability of RNN models in classifying valence states from multimodal physiological data?
(2) How do different architectures and hyper-parameters of RNN models affect classification performance?
Edge Device Deployment
(3) How can the models be optimised to run efficiently on edge devices without compromising classification accuracy?

1.4 Contribution

In this work, multimodal physiological data from cyclists—heart-rate variability (HRV) and electrodermal activity (EDA) in particular—are used to investigate the applicability of Recurrent Neural Networks (RNN) to emotion recognition. Prediction models that can be deployed on edge devices attached to bicycles are created using these signals. With this approach, emotion identification systems under dynamic cycle situations should be more capable and adaptive.

The study's novel approach to emotion identification in cycling involves combining real-time physiological data gathering with advanced deep learning algorithms. This makes it innovative and significant. In contrast to previous research that focuses on non-cycling activities or static locations, this study tackles the particular difficulties associated with emotion identification in dynamic cycling situations.

2 RELATED WORK

This section provides a literature review on multiple topics, such as emotion recognition by physiological data and applications of it in deep learning, to create the base for the next steps needed in this research.

2.1 Emotion Recognition via Physiological Data

Recent studies have successfully employed wearable devices such as wristbands and chest belts to monitor emotional responses. These devices measure physiological signals like HRV and EDA, which are indicative of emotional states. For example, [17] demonstrated the use of HRV and EDA data to recognise stress and anxiety levels in real time. These findings suggest that physiological data can be a reliable indicator of emotional states. However, the practical application of these technologies in enhancing cyclist safety remains an under explored area. Another study [26] validates the use of physiological signals, such as HRV and electrodermal activity EDA achieving a relatively high accuracy rate for emotion recognition. This study's recommendation for improving the overall performance is to explore other deep learning approaches, such as neural networks, to compare to their approach.
A comprehensive discussion of physiological signal-based emotion recognition was given in [21]. A number of key insights were emphasised, including the importance of feature extraction techniques, and the necessity of multimodal data fusion for improving accuracy. For instance, their review stressed that integrating features from several signals (e.g, HRV,EDA,EEG) improves the robustness of models used for emotion recognition. This is relevant for this research, as the goal is to include features derives from multiple physiological signals.

2.2 Applications of Deep Learning in Emotion Recognition

Deep learning models, particularly those designed for time-series data, have shown significant potential in emotion recognition tasks. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are especially effective for handling sequential data due to their ability to retain information over time. Studies [16, 28], have demonstrated that LSTM and GRU models can accurately predict emotional states from physiological data streams. Despite these advancements, there is limited research [17, 26, 28], on applying these models in real time, naturalistic environments such as cycling. Addressing this gap is crucial for developing interventions that enhance safety through immediate valence state recognition.

2.3 Wearable Technology and Real Time Data Processing

Wearable technology has advanced to the point where devices like the Empatica E4 wristband can provide continuous, real time physiological data. Research [18] has explored the use of these devices in various contexts, such as monitoring stress in workplace environments and tracking emotional responses in social settings. highlight the potential of wearables to offer detailed, real time insights into valence states, making them suitable for applications in dynamic environments like cycling.

2.4 Feasibility of Edge Computing for real time Applications

Deploying deep learning models on edge devices for real time applications presents several challenges, including computational constraints, power consumption, and data privacy. Studies like [30] have addressed some of these challenges by optimising model architectures to run efficiently on edge devices. For instance, they implemented energy-efficient algorithms that significantly reduce the computational load without compromising accuracy. Additionally, [1] has emphasised the importance of local data processing to ensure privacy and reduce latency, making real time interventions more feasible. These insights will inform our approach to developing models that are accurate, efficient, and practical for real time use on bike-mounted edge devices.

3 METHODS OF RESEARCH

3.1 Data Handling and Processing

The data for this study was pre-collected using Empatica E4 wristbands. Each participant's data was organised in individual folders containing CSV files with various physiological measurements, including Blood Volume Pulse (BVP) and electrodermal activity (EDA). Additionally, each participant's bike was equipped with three buttons that they could press to log their valence state (negative, neutral, and positive). The button press data, including timestamps, was also stored in CSV files within each participant's folder. Initially, the data collected was from 30 participants. However, several data quality issues led to the exclusion of some datasets. Some participants were missing either the Empatica files, and others had errors with button presses. For example, certain participants recorded an unrealistic number of button presses (e.g., 250 button touches during a single ride), making it impossible to correlate with physiological data. Additionally, some participants had button presses recorded before or after their rides, or only on the third button, resulting in an imbalance in the dataset. To maintain a balanced dataset, the problematic records were excluded, Upon receiving the remaining pre-collected data, the initial step involved organising and preparing the data for analysis. Python scripts were employed to read and structure the data, ensuring that all datasets were correctly aligned by timestamps and participant IDs. This step was crucial for maintaining the integrity of the measurements across different devices and sessions. Initially, an attempt was made to manually clean and preprocess the data. However, the theoretical knowledge of physiological data was insufficient for this task. Recognising this gap, various libraries were explored for processing and feature generation of physiological data. After multiple iterations were tested the library ultimately used was the FLIRT library, which is specifically tailored for handling data from Empatica devices. The FLIRT library preprocesses the data according to literature recommendations, ensuring that the preprocessing steps align with established research standards. This library handles missing values, noise, and outliers in the data, providing a clean and reliable dataset for further analysis. Additionally, FLIRT calculates 178 features related to HRV, EDA, and accelerometer data (ACC), encompassing statistical measures and patterns relevant to emotion recognition. After using the FLIRT library, additional post-processing steps were necessary to handle any remaining infinite (Inf) and not-a-number (NaN) values. Python scripts were used to identify and remove these values, ensuring the dataset was suitable for model training. The cleaned data was then saved back into individual participant files for further use.

3.2 Model Development

The development of the predictive models was carried out using TensorFlow, a robust deep learning framework well-suited for handling time-series data. TensorFlow's extensive support for sequential data processing and its powerful library of tools and resources made it the ideal choice for this research. The model architecture focused on two types of Recurrent Neural Networks (RNNs): Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) networks. These architectures were selected for their ability to retain information over time, which is essential for analysing sequential physiological data.

3.2.1 *Model Architecture.*

The models were then designed with the first two LSTM/GRU layers, each followed by a dropout layer to prevent overfitting. The first layer had 128 units, and the dropout rate 0.2, and the second layer had 64 units and a dropout rate of 0.1 After, a dense layer with 16 units and ReLU activation was added, followed by an output layer with softmax activation to classify the emotional states. The model was compiled with the Adam optimiser and a learning rate of 0.001, using sparse categorical cross entropy as the loss function. Early stopping was implemented to monitor the validation loss and restore the best weights if the validation loss did not improve for five consecutive epochs.

3.2.2 *Hyperparameter Tuning.*

The hyperparameters were fine-tuned through a manual grid search process.
This involved varying units, drop rates, learning rates, batch sizes, and optimisers.
The best parameters identified were: 128 units, a dropout rate of 0.2, a learning rate of 0.005, a batch size of 64, and the Adam optimiser. These parameters ensured the model's optimal performance

3.3 Model Training and Validation

Training the models required labelled data to teach the RNNs to recognise emotional states. The labelled data comprised button press events, where users self-logged their valence state(negative, neutral, and positive). In training models, 50 epochs were used with the 'early stopping' technique to minimise the overfitting of the model. The training was also stopped early at the point where the validation loss failed to drop in five consecutive epochs. In order to cross-reference the button presses with the physiological measurements, a time window beginning 60 seconds prior to a button press and ending 60 seconds after the button press was correlated, creating sequences with this data. The data was then made into sequences to match the input structure of the LSTM Model. The performance of the demonstrated models was assessed considering accuracy, precision, recall, F1-score, and AUC-ROC coefficients. Confusion matrix and classification reports enabled the exploration of the specific attributes of the models' classification capacity to select the most appropriate valence state. The plot of training and validation accuracy or loss was used to show the learning characteristics of the models.

3.4 Optimisation for Edge Devices

Because of their limited processing resources and energy restrictions, it is essential that the models produced have the ability to operate properly on edge devices installed on bicycles. Since many bike-mounted computers lack specific specifications, these constraints make it difficult to generalise the criteria for optimisation.

The study took into account the usual limitations of low-power, ARM Cortex-based processors—which are frequently employed in micro-controller applications and other related embedded systems—in order to tackle this issue. Because of their low power consumption and real-time processing capabilities, these processors are a good starting point for the present research. Their RAM(Random

Access Memory) ranges from 64MB to 2GB, their clock speeds are often lower, and their power consumption is typically limited to less than 5 watts [2, 29].

Several optimisation methods were used to make sure the model could potentially fit within the limitations of bike-mounted devices. The parameters for data processing were first adjusted. Typically, the FLIRT library preprocesses data with a time step of one second and a window length of 180 seconds. These parameters were changed to a 60-second window length and a 10-second time step, which greatly decreased the data amount without sacrificing important information. This change of reducing the window time, means around ten times less data, which was easier to handle for edge device processing. Another crucial step in the optimisation process was feature selection. The initial 178 features have been reduced back to a more manageable subset due to the resource constraints of edge devices. Correlation tests, Recursive Feature Elimination (RFE), and Random Forest (RF) algorithms were used in the feature selection process in order to assess each feature's significance. After much consideration, about thirty features that were both computationally efficient and offered significant insights into emotional states were chosen. These selected features were then used to train the models, ensuring that the data input size was optimised for edge device constraints without compromising the accuracy of emotion recognition. Additionally, in order to decrease the size of the model and increase inference efficiency, model optimisation methods including quantisation and pruning were investigated. By transforming the model weights and activation's from 32-bit floating-point to 8-bit integers by quantisation, the model's size and computing load were greatly decreased. Conversely, pruning reduced model complexity without appreciably sacrificing performance by eliminating unnecessary or insignificant connections from the neural network [10, 14].

TensorFlow's profiling tools were used to quantify the inference time, processing power, and energy consumption on a development computer in order to assess the performance of the optimised model [25].

4 RESULTS

This section presents the performance of the models, supported by figures and graphs to illustrate the process for the training, the validation, and the overall accuracy of the models.

4.1 Model Performance Metrics

The performance of the LSTM and GRU models was evaluated on the test dataset. Here, is presented the accuracy, precision, recall, and F1-score for each model. Additionally, the performance is visualised through training and validation curves as well as confusion matrices. For the LSTM model, the accuracy was 0.80, the precision was 0.7997, recall was 0.80, F!-score was 0.7995, and AUC-ROC was 0.9358. The detailed classification report showed precision, recall, and F1-scores for the negative (Class 0), neutral (Class 1), and positive (Class 2) classes. The positive class had the highest scores, indicating the model's robustness in detecting positive emotional states. The LSTM model's classification report is as follows: Class 0: Precision 0.74, Recall 0.78, F1-score 0.76 Class 1: Precision 0.73, Recall 0.68, F1-score 0.71 Class 2: Precision 0.86, Recall 0.86, F1-score 0.86 Similarly, the GRU model achieved an accuracy of 0.7947, precision of 0.7923, recall of 0.7947, F1-score of 0.7932, and AUC-ROC of 0.9270. The GRU model's classification report showed strong performance, with slightly lower precision and recall for the neutral class compared to the LSTM model. The GRU model's classification report is as follows: Class 0: Precision 0.78, Recall 0.77, F1-score 0.78 Class 1: Precision 0.70, Recall 0.65, F1-score 0.68 Class 2: Precision 0.84, Recall 0.87, F1-score 0.86

4.2 Cross-Validation Results

To ensure the robustness and generalizability of the models, k-fold cross-validation was employed. The cross-validation performance metrics demonstrated the stability of both the LSTM and GRU models. The LSTM model achieved a final epoch validation accuracy of 0.8309 ± 0.0061 and a final epoch validation loss of 0.3990 ± 0.0231. The GRU model had a final epoch validation accuracy of 0.8298 ± 0.0134 and a final epoch validation loss of 0.4112 ± 0.0401. These results indicate that both models are stable and reliable.

4.3 Figures and Tables

The training and validation accuracy and loss curves illustrate how the models learned over time and indicate when early stopping was triggered. Figures 2 and 3 show the training and validation accuracy and loss for the LSTM model, respectively. Similarly, Figures 4 and 5 present the training and validation accuracy and loss for the GRU model. These plots help in understanding the learning dynamics of the models. Confusion matrices provide a detailed view of the models' performance by showing the actual vs. predicted classifications. Figures 6 and 7 depict the confusion matrices for the LSTM and GRU models, respectively. These matrices highlight the classification accuracy for each emotional state, indicating areas where the models performed well and where improvements might be needed. Detailed classification reports for each model summarise the precision, recall, and F1-scores for each class. Tables 1 and2 present the classification reports for the LSTM and GRU models, respectively. These reports provide a comprehensive overview of the models' classification performance. Figure 8, 9 shows the k-fold cross-validation mean and variance accuracy and loss the LSTM model, highlighting the consistency and variance of model performance across different folds. Figures 10,11, 12 show the accuracy and loss per epoch for both the LSTM and GRU models without the mean and variance, providing insight into the model performance over each training epoch. Table 3 shows the inference time, memory usage during inference, and model size for both the LSTM and GRU models in their baseline, pruned, and quantised versions.

Table 1. LSTM Model Classification Report

Class	Precision	Recall	F1-Score	Support
0	0.74	0.78	0.76	145
1	0.73	0.68	0.71	133
2	0.86	0.86	0.86	287
Accuracy		0.80		
Macro Avg		0.78	0.78	0.78
Weighted Avg		0.80	0.80	0.80

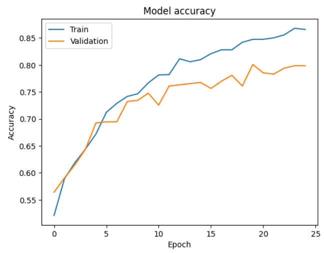

Fig. 2. LSTM model accuracy

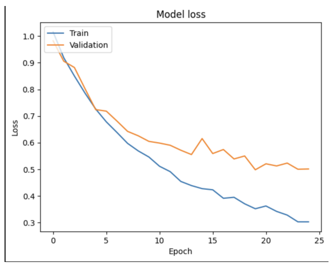

Fig. 3. LSTM model loss

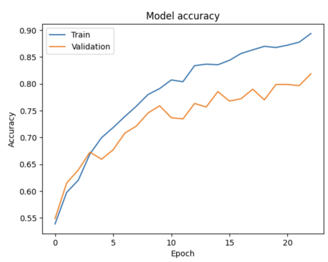

Fig. 4. GRU model accuracy

Fig. 5. GRU model loss

4.4 Summary of Key Metrics

In summary, both the LSTM and GRU models showed strong performance in recognising emotional states from wearable data. The LSTM model achieved a higher consistency in performance, while the GRU model demonstrated slightly more variance but comparable accuracy. Cross-validation confirmed the robustness of the models, highlighting their potential for real time emotion recognition applications.

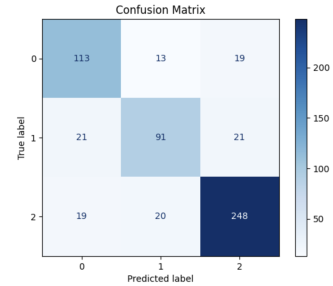

Fig. 6. LSTM confusion matrix

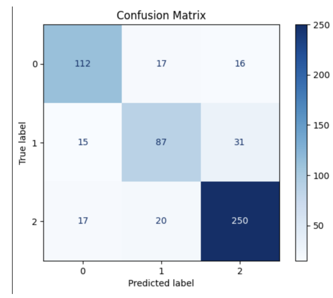

Fig. 7. GRU confusion matrix

Table 2. GRU Model Classification Report

Class	Precision	Recall	F1-Score	Support
0	0.78	0.77	0.78	145
1	0.70	0.65	0.68	133
2	0.84	0.87	0.86	287
Accuracy		0.79		
Macro Avg		0.77	0.77	0.77
Weighted Avg		0.79	0.79	0.79

5 DISCUSSION

5.1 Introduction

The purpose of this study was 3to assess how well multimodal physiological data from cyclists might be used to characterise their emotional states using Recurrent Neural Network (RNN) models, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. The main research question posed was: **How effectively can RNN models classify cyclists' valence states using multimodal physiological data for potential deployment on bike-mounted edge devices?** Furthermore, the research investigated the sub-questions in regard to Model Performance and Edge Device Deployment:

Model performance First, the study discovered that both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures performed well in predicting valence states from HRV and EDA data, indicating the accuracy and dependability of RNN models. The models' capacity to reliably identify emotional states from physiological inputs is shown by their satisfactory accuracy

Model	Inference Time (s)	Memory Usage (MiB)	Model Size (MB)
LSTM			
Baseline	0.043687	1.28125	1.59
Pruned	0.044265	2.765625	0.54
Quantized	0.000479	0.390625	0.15
GRU			
Baseline	0.044456	3.765625	0.39
Pruned	0.042778	1.65625	0.05
Quantized	0.000438	0.1875	0.05

Table 3. Inference time, memory usage, and model size for LSTM and GRU models in their baseline, pruned, and quantized versions.

rates. This result is noteworthy because it validates these models' potential for real-world bicycle safety applications. The study also looked at how various designs and hyper-parameters affected the performance of the model. It was found that the LSTM networks perform the same overall as GRU networks. Enhancing classification accuracy also required careful adjustment of hyper-parameters like the number of layers and units in the RNNs. The hyperparameters were selected after a grid-search, because at the start of the experiment there were cases when the model was not learning at all, or the loss was not decreasing at all through the epochs. The outcomes highlight how crucial it is to optimise these parameters in order to get the most performance out of the models.

Edge Device Deployment The initial, unoptimised models required more processing power and memory than the typical edge devices employed in this study, making them too resource-intensive. Several optimisation strategies were used to address these constraints.

One of the main methods for lowering the model's size and processing load was quantization. The models were much more efficient when the weights and activation's were changed from 32-bit floating-point numbers to 8-bit integers. Because of this decrease in numerical precision, the models were nevertheless able to function accurately even with the constrained processing capacity of edge devices. As a result, quantization made sure that the models could identify emotions in real time without straining the computing power of the device. Pruning was used in addition to quantization to further optimise the models. This method reduced the complexity of the model by removing less important neurons and connections from the neural network. Pruning allowed the models to function more effectively by reducing memory consumption and computational demands. The models maintained their accuracy in spite of this reduction in complexity, proving that the optimisations had no adverse effect on their capacity to accurately categorise emotions.

The implementation of the RNN models on the edge devices mounted on bikes necessitates the use of these optimisation techniques. They demonstrated that real-time emotion identification on resource-constrained platforms is feasible without needing a compromise between computational efficiency and classification accuracy, as the accuracy remained on par with previous levels.

5.2 Limitations

The study faced several limitations, including the quality and completeness of the data. Some participants had missing or noisy data, which required extensive preprocessing. Additionally, the dataset size was relatively small, and further validation with larger datasets is necessary to confirm the models' generalizability. The potential for overfitting was mitigated using early stopping, but more robust methods could be explored in future research. Environmental factors were an additional limitation of the study, as it omitted to take into consideration elements that could potentially impact the physiological responses and emotional states of cyclists, such as weather patterns and traffic density. The results' potential for generalisation may be impacted by this disregard for environmental influences. Moreover, no real-world testing was done on edge devices installed on bikes, even though the models were evaluated in controlled settings. Therefore, further validation is required to guarantee the models' robustness in a variety of dynamic cycling scenarios.

5.3 Future Work

Future research should explore additional physiological signals, such as respiratory rate or skin temperature, to enhance emotion recognition accuracy. Testing the models on more diverse and larger datasets would provide better validation of the findings. Additionally, integrating other machine learning techniques, such as ensemble learning, could further improve model performance. Developing lightweight models optimised for real time processing on wearable devices remains a key area for future investigation. For bike manufacturers considering the integration of emotion detection systems, it is recommended that each individual participant manually label their emotional state when they first get a new e-bike. This personalised labelling should continue until the local model learns from their specific data. Since each individual has different baselines and emotional responses, this approach ensures a more accurate and personalised emotion recognition system, rather than relying on data from other participants.

6 CONCLUSION

6.1 Summary of Key Findings

This study successfully demonstrated the use of LSTM and GRU models for emotion recognition from physiological data collected via wearables. Both models achieved high accuracy, validating the potential of RNNs in real time emotion detection.

6.2 Significance of the Study

The research contributes to the field of emotion recognition by showcasing the effectiveness of advanced deep-learning models. The findings highlight the potential applications of real time emotion monitoring, paving the way for innovative uses in mental health, personalised user experiences, and adaptive interfaces.

6.3 Final Remarks

In conclusion, this study underscores the feasibility and effectiveness of using wearables for real time emotion recognition. Future research should aim to build on these findings, exploring new signals and techniques to further enhance the capabilities and applications of emotion recognition systems. The integration of such technology holds promise for significantly improving user experience and well-being across various domains.

REFERENCES

[1] 2021. Designing an affective states input tool for biking. https://doi.org/10.1109/XXXXXX

[2] ARM. 2023. *Arm M4 chip.* https://developer.arm.com/Processors/Cortex-M4 Accessed: 2024-06-27.

[3] S. Brink. 2022. De meefietslijn. (2022). https://www.fietsersbond.nl/nieuws/de-meefietslijn-2/ Accessed: 2022-10-11.

[4] K. Buchholz. 2024. Where cyclists are going places. (2024). https://www.statista.com/chart/25156/share-using-bike-for-transportation-regularly/ Accessed: 2024-02-22.

[5] Ralph Buehler and John Pucher. 2021. *Cycling for Sustainable Cities.* MIT Press.

[6] S. F"oll, M. Maritsch, F. Spinola, V. Mishra, F. Barata, T. Kowatsch, E. Fleisch, and F. Wortmann. 2021. FLIRT: A feature generation toolkit for wearable data. *Computer Methods and Programs in Biomedicine* 212 (2021), 106461. https://doi.org/10.1016/j.cmpb.2021.106461

[7] Lukas Giessen, Alex Zhang, Qing Lyu, Amelia Tan, and Alexander Z. Wang. 2022. Multi-scale Transformer-based Network for Emotion Recognition from Multi Physiological Signals. *arXiv preprint arXiv:2203.06935* (2022). https://arxiv.org/abs/2203.06935

[8] E. Gorgul, L. Zhang, F. G"unther, and C. Chen. 2019. Mapping human response to street experience. (2019). https://doi.org/10.1145/3341162.3343793 Accessed: 2019-09-09.

[9] Mutaza Haider. 2023. The Dutch Disease. https://www.linkedin.com/pulse/dutch-disease-cyclists-account-40-road-deaths-murtaza-haider-hbetc Accessed: n.d..

[10] Song Han, Huizi Mao, and William J Dally. 2015. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149* (2015).

[11] J. Joshi, K. Wang, and Y. Cho. 2023. PhysioKit: an Open-Source, Low-Cost physiological computing toolkit for Single- and Multi-User studies. *Sensors* 23 (2023), 8244. https://doi.org/10.3390/s23198244

[12] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Sch"olzel, and S. H. A. Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[13] Amy S. Martin, Michael Kilpatrick, Maria Ashe, and Rory Guerin. 2020. The Impact of Stress and Anxiety on Cycling Safety. *Safety Science* 128 (2020), 104744.

[14] Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, and Jan Kautz. 2016. Pruning Convolutional Neural Networks for Resource Efficient Inference. *arXiv preprint arXiv:1611.06440* (2016).

[15] NOS. 2022. Steeds meer fietsongelukken met ernstig letsel tot gevolg. (2022). https://nos.nl/artikel/2424819-steeds-meer-fietsongelukken-met-ernstig-letsel-tot-gevolg Accessed: 2022-04-11.

[16] D. Pollreisz and N. Taherinejad. 2017. A simple algorithm for emotion recognition, using physiological signals of a smart watch. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2353–2356. https://doi.org/10.1109/EMBC.2017.8037328

[17] M. Ragot, N. Martin, S. Em, N. Pallamin, and J. Diverrez. 2017. Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors. In *Advances in intelligent systems and computing.* 15–22. https://doi.org/10.1007/978-3-319-60639-2_2

[18] L. Santamaria-Granados, J. Mendoza-Moreno, Á. Astaiza, M. Organero, and G. Ramírez-González. 2021. Tourist Experiences Recommender System Based on Emotion Recognition with Wearable Data. *Sensors (Basel, Switzerland)* 21 (2021). https://doi.org/10.3390/s21237854

[19] A. Scarano, M. Aria, F. Mauriello, M. R. Riccardi, and A. Montella. 2023. Systematic literature review of 10 years of cyclist safety research. *Accident Analysis and Prevention* 184 (2023), 106996. https://doi.org/10.1016/j.aap.2023.106996

[20] P. Schmidt, A. Reiss, R. D"urichen, and K. Van Laerhoven. 2019. Wearable-Based Affect Recognition—A review. *Sensors* 19, 19 (2019), 4079. https://doi.org/10.3390/s19194079

[21] Lin Shu, Jinxuan Xie, Meng Yang, Zheng Li, Zhenyu Li, Dong Liao, Xiyang Xu, and Xingshuo Yang. 2018. A Review of Emotion Recognition Using Physiological Signals. *Sensors* 18, 7 (2018), 2074. https://doi.org/10.3390/s18072074

[22] MIT Sloan. 2023. *Emotion AI, explained.* https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained Accessed: 2024-06-27.

[23] Statista. 2023. Number of cyclist road fatalities in the Netherlands 1996-2022. (2023). https://www.statista.com/statistics/523310/netherlands-number-of-cyclist-road-fatalities/ Accessed: 2023-04-18.

[24] Statista. 2023. Weekly cycling frequency around the world 2022, by country. (2023). https://www.statista.com/statistics/1353329/cycling-frequency-worldwide/ Accessed: 2023-12-18.

[25] TensorFlow. 2023. *TensorFlow Profiler.* https://www.tensorflow.org/guide/profiler Accessed: 2024-06-27.

[26] N. Thammasan. 2018. Practical Emotion Recognition using Wearable Brain and Physiological Sensors. https://doi.org/10.18910/69712

[27] NL Times. 2023. Over 25% more traffic deaths last year; Many older cyclists killed. (2023). https://nltimes.nl/2023/04/18/25-traffic-deaths-last-year-many-older-cyclists-killed Accessed: 2023-04-18.

[28] Z. Wang, Z. Yu, B. Zhao, B. Guo, C. Chen, and Z. Yu. 2020. EmotionSense: An Adaptive Emotion Recognition System Based on Wearable Smart Devices. *ACM Trans. Comput. Heal.* 1 (2020), 20:1–20:17. https://doi.org/10.1145/3384394

[29] Joseph Yiu. 2013. *Embedded Systems Design with ARM Cortex-M Microcontrollers.* Elsevier.

[30] T. Zhang, A. Ali, C. Wang, A. Hanjalic, and P. César. 2020. CorrNet: Fine-Grained Emotion Recognition for Video Watching Using Wearable Physiological Sensors. *Sensors (Basel, Switzerland)* 21 (2020). https://doi.org/10.3390/s21010052

A APPENDIX

A.1 AI Statement

During the preparation of this work the author used ChatGPT to better organise the reference list, to search for sources and provide guidelines for feedback. Quillbot and Microfost built-in editor were also used to ensure grammatical correctness and improve clarity. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the work.

A.2 Additional Figures

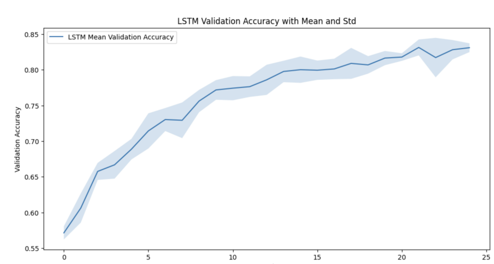

Fig. 8. LSTM accuracy mean and variance on k-fold

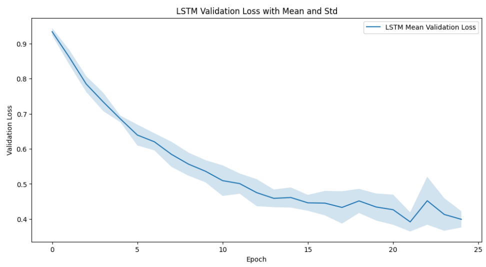

Fig. 9. Lstm loss mean and variance on k-fold

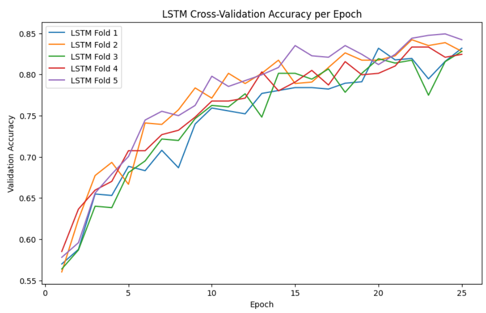

Fig. 10. LSTM Cross-Validation loss per Epoch

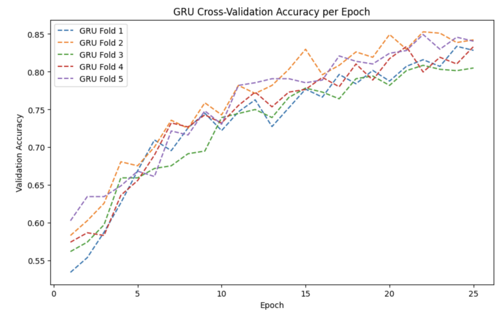

Fig. 11. GRU Cross-Validation Accuracy per Epoch

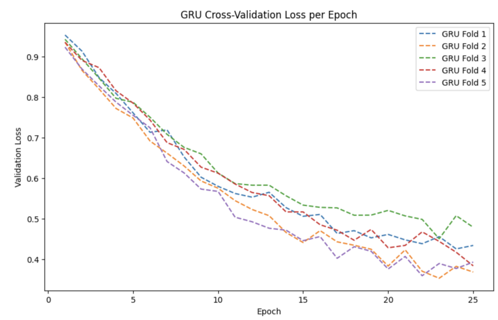

Fig. 12. GRU Cross-Validation loss per Epoch