Advanced Driving Style Classification using Temporal Convolutional Networks on Vehicle Data

Rima ALAYA[1*], Faouzi JAIDI[2] and Aymen TOUHENT[3]

[1*]National School of Engineers of Carthage, University of Carthage, Tunis, Tunisia.
[2]Higher School of Communication of Tunis (Sup'Com), Digital Security Research Lab, University of Carthage, Tunis, Tunisia.
[3]XELERO, Tunis, Tunisia.

*Corresponding author(s). E-mail(s): rima.alaya@enicar.ucar.tn;
Contributing authors: faouzi.jadi@supcom.tn; aymen@xelero.io;

Abstract

This study addresses the crucial task of driving style classification within Intelligent Transport Systems (ITS). We introduce a robust methodology utilizing an advanced data preprocessing and feature engineering pipeline applied to the public Mafalda dataset. Our primary focus is on the performance of a Temporal Convolutional Network (TCN), benchmarked against a Random Forest (RF) and a standard Artificial Neural Network (ANN). The results demonstrate exceptional performance, with all models achieving accuracy, precision, recall, and F1-scores consistently above 97%. Notably, the TCN exhibited superior capability in modeling the temporal dependencies of the data, achieving a perfect recall for aggressive driving detection. These findings represent a substantial improvement over existing state-of-the-art results, reinforcing the TCN's effectiveness for dynamic driving behavior analysis and paving the way for more reliable and safer automotive AI systems.

Keywords: Driving Style Classification, OBD-II, Temporal Convolutional Network (TCN), Machine Learning, Random Forest, Artificial Neural Network, Mafalda Dataset

1 Introduction

Artificial Intelligence (AI) is rapidly transforming various industries, with its influence on the automotive sector being particularly significant. A core aspect of this change is the detailed analysis of driver behavior, which is essential for advancing Intelligent Transport Systems (ITS). The ability to accurately classify driving styles is crucial for a variety of applications, including personalized insurance, road safety enhancements, and eco-driving. A primary challenge in this area is the precise identification of different driving styles—such as aggressive, normal, or even pace—from the complex and inherently temporal data streams generated by On-Board Diagnostics (OBD-II) systems. These data, which include parameters like engine RPM and vehicle speed, are complicated by their sequential nature, necessitating robust computational models that can effectively capture dynamic patterns for reliable classification. This research is a foundational element in creating an intelligent AI engine for the automotive industry, where accurate driving style classification is a vital module for more comprehensive vehicle intelligence systems.

The remainder of this article is structured as follows: Section 2 provides essential background and a review of related work, along with the research contributions. Section 3 delineates the comprehensive methodology. Section 4 presents the experimental results. Section 5 offers a thorough discussion of these results, and Section 6 concludes the article.

2 Research Context and Contributions

While driving style classification remains an active and evolving area of research, a persistent need exists for the development of more robust, precise, and generalizable models. Existing approaches often struggle to fully exploit the rich temporal dependencies embedded within OBD-II data, which are critical for nuanced behavioral analysis. This study addresses these limitations by offering several significant contributions:

- This work involves the development and rigorous evaluation of advanced machine learning models, including Random Forest, a standard Artificial Neural Network, and a Temporal Convolutional Network, specifically tailored for precise driving style classification using the public Mafalda dataset [1].
- Significantly improved performance is demonstrated across key metrics (accuracy, precision, recall, and F1-score) compared to recent state-of-the-art approaches. A direct quantitative comparison is provided against the results reported by Al-refai et al. (2024) [2], highlighting the advancements achieved.
- Through meticulous analysis, the study identifies key features within OBD-II data that exhibit the strongest correlation with distinct driving styles, thereby providing valuable insights into the underlying kinematic profiles that characterize different behaviors.
- Empirical evidence underscores the superior effectiveness of Temporal Convolutional Networks (TCNs) in capturing complex temporal patterns inherent in sequential

driving data, establishing their suitability for this challenging task. The sophisticated architecture of the TCN, including elements like spatial dropout, contributes to its robustness and ability to generalize effectively to unseen data.

2.1 OBD-II Data and the Mafalda Dataset

On-Board Diagnostics II (OBD-II) systems are standard interfaces that provide real-time access to a wide array of vehicle parameters, including Revolutions Per Minute (RPM), vehicle speed, engine temperature, and fuel consumption [3]. These systems are vital for modern vehicle diagnostics and have become a crucial source of data for advanced analysis of driver behavior. For this study, the Mafalda dataset [1] was used as a key public resource. This dataset is characterized by its time-series nature and its detailed annotations of various driving styles, which are essential for ensuring research reproducibility.

2.2 Existing Machine Learning Approaches for Driving Style Classification

The field of driving style classification has seen a wide exploration of various machine learning approaches. These range from traditional methods like Support Vector Machines (SVMs) [4] and ensemble techniques such as Random Forests [5] to different types of Artificial Neural Networks.

Within this context, Conti et al. [6] proposed an SVM-based approach specifically for aggressive driving pattern recognition using a private dataset.More sophisticated architectures designed to capture temporal dependencies have also been explored. Khan and Lee [7] developed a hybrid model combining a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network, tested on the public UAH-Drive dataset [8] to classify behaviors into categories such as normal and aggressive. Similarly, Wang and Chen [9] utilized a simple LSTM network to analyze driving styles from telematics data using a private dataset, further demonstrating the efficacy of temporal models.

Despite these advances, persistent challenges remain. A critical issue is the efficient processing and full utilization of the inherent temporal dependencies within sequential OBD-II data. Many conventional models may not adequately capture the dynamic evolution of driving behaviors over time, which is essential for accurate classification.

2.3 Review of Al-refai et al. (2024)'s Work

The study by Al-refai et al. (2024) [2] is a highly relevant benchmark for the current research, particularly because it uses the same Mafalda dataset. Their model utilized 14 input features from both OBD and smartphone sensors. The machine learning models implemented included Feed-forward Artificial Neural Networks (ANNs) and Bagging ANNs. Their best reported results for driving style classification are summarized in Table 1, showing a best F1-score of 0.85. While these results are commendable, their reliance on conventional feed-forward ANNs may not have fully leveraged the complex temporal patterns found in sequential OBD-II data. This limitation is particularly relevant as the current study focuses on demonstrating superior performance using OBD-II data.

Table 1: Results of driving-style classification by Al-refai et al. (2024) [2].

Model	Precision	Recall	Accuracy	F1 Score
Proposed ANN	0.86	0.81	0.92	0.83
Proposed ANN + dropout	0.80	0.88	0.92	0.84
Proposed ANN + bagging	0.79	0.91	0.92	0.85
SVM (reference)	0.68	0.84	0.82	0.71

3 Proposed Approach

3.1 Data Acquisition and Preprocessing

The Mafalda dataset [1] served as the exclusive data source. A multi-stage, robust preprocessing pipeline was meticulously applied. The steps included:

- **Advanced Cleaning:** This involved comprehensive handling of vehicle stops (defined as speeds ≤ 1.0 m/s), and the management of null values through intelligent imputation. Linear interpolation was applied for continuous variables like speed, while forward-fill (ffill) and backward-fill (bfill) methods were used for other parameters. Outlier detection and correction were performed using the Interquartile Range (IQR) method.
- **Physical Data Validation:** To maintain realism, values were clipped to realistic physical bounds, ensuring that all data points adhered to plausible vehicle operating conditions.
- **Trip Segmentation:** The continuous data streams were segmented into distinct 'trips' by detecting prolonged vehicle stops.

3.2 Feature Engineering and Selection

A sophisticated feature engineering process was undertaken. This meticulous approach is a major contributing factor to the enhanced performance observed. The derived features included:

- **Contextual Features:** Binary indicators were created to identify specific driving phases, such as 'is_accelerating', 'is_braking', 'is_cruising', and 'is_stopped'.
- **Temporal Features (per trip):** Rolling statistics, including mean and standard deviation, were calculated for key variables like speed and acceleration over predefined time windows.
- **Aggressive Indicators:** Features were derived to detect abrupt events, such as 'Harsh Acceleration' and 'Harsh Braking'.

- **Statistical Moments (per trip):** Higher-order statistical moments, specifically skewness and kurtosis, were computed for critical variables to capture the shape and distribution of data patterns.

Furthermore, data were aggregated into 5-second temporal windows. A thorough correlation analysis was performed, consistently revealing the dominance of speed-related characteristics as primary indicators of driving style, as shown in Table 2.

Table 2: Top 15 Features Correlated with Driving Style.

Feature Name	Correlation
VehicleSpeedInstantaneous	0.149
VehicleSpeedVariance	0.138
VehicleSpeedAverage	0.115
IntakeAirTemperature	0.108
EngineLoad	0.082
VerticalAcceleration	0.075
EngineCoolantTemperature	0.069
MassAirFlow	0.031
AltitudeVariation	0.028
VehicleSpeedVariation	0.020
FuelConsumptionAverage	0.019
EngineRPM	0.015
LongitudinalAcceleration	0.008
ManifoldAbsolutePressure	0.004

3.3 Model Architectures

This research is focused primarily on the Temporal Convolutional Network (TCN) because of its remarkable ability to handle sequential data. The Random Forest (RF) and a standard Artificial Neural Network (ANN) were also examined and tested on the same carefully preprocessed datasets to provide a comparative study and ultimately a baseline for existing state-of-the-art models.

3.3.1 Temporal Convolutional Network (TCN)

The Temporal Convolutional Network (TCN) was selected as the primary model due to its demonstrated excellence in processing time-series data, making it particularly well-suited for the sequential nature of OBD-II streams. The TCN's architecture is specifically designed to capture long-range temporal dependencies through the use of dilated causal convolutions.

Key architectural elements of the TCN included:

- An input layer equipped with batch normalization to stabilize training.
- A stack of deep TCN blocks, each meticulously designed to incorporate dilated causal convolutions, further batch normalization layers, ReLU activations [10], and

spatial dropout layers. The use of spatial dropout, a specialized regularization technique for convolutional layers where entire feature maps are dropped, significantly contributes to the model's robustness and generalization, especially when dealing with sequential data, by preserving spatial correlations while preventing overfitting.

- The integration of residual connections within each block and skip connections between blocks was crucial for facilitating gradient flow and enabling the learning of increasingly complex representations across layers.
- An attention mechanism was incorporated to dynamically weight the importance of different temporal features, allowing the model to focus on the most relevant parts of the driving sequence.
- An enhanced global pooling layer, combining both average and maximum pooling, was used to effectively aggregate sequential information into a fixed-size representation suitable for classification.
- Finally, dense classification layers with additional regularization and dropout led to a softmax output layer, producing the final prediction of driving styles.

This advanced architecture allows the TCN to far exceed the performance of conventional models in complex sequential data analysis, and it is especially effective in the case of driving style classification.

The conceptual architecture of the enhanced TCN is presented in Fig. 1.

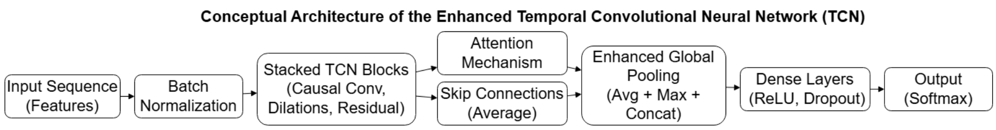


Fig. 1: Conceptual architecture of the enhanced Temporal Convolutional Network (TCN) model.

3.3.2 Comparison Models: Random Forest (RF) and Artificial Neural Network (ANN)

Random Forest (RF): Random Forest was implemented as a robust ensemble model [5]. Its capacity to maintain non-linear operational relationships, along with a resilience to overfitting, makes it a very effective model for performing with complex datasets. In this study, this model provided a strong, reliable baseline for comparison against the more advanced neural network architectures, an indication of the performance that a well-established ensemble method demonstrated on the preprocessed data.

Artificial Neural Network (ANN): A standard Artificial Neural Network (ANN) was implemented to represent a more conventional deep learning approach [11]. Its architecture comprised multiple hidden layers utilizing Rectified Linear Unit (ReLU) activation functions [10]. The model was optimized using the Adam optimizer [12], and dropout layers were incorporated as a regularization technique to prevent overfitting and enhance generalization capabilities. This ANN serves as a direct comparison to the TCN, highlighting the benefits of specialized temporal architectures over general-purpose deep learning models for sequential data.

3.4 Experimental Setup and Evaluation Metrics

For robust model evaluation, the dataset was partitioned into three distinct sets: 70% for training, 15% for validation, and 15% for testing. This division ensures that the models are trained on a substantial portion of the data, their performance is monitored during training, and their final generalization capability is assessed on unseen data [13].

The models were rigorously evaluated using the following standard metrics, crucial for assessing classification performance[14]:

1. **Overall Accuracy (Accuracy):** Represents the proportion of correctly classified instances, as defined in Equation 1.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2. **Precision:** Indicates the proportion of true positive predictions among all positive predictions, calculated as shown in Equation 2.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

3. **Recall:** Represents the proportion of true positive predictions among all actual positive instances, as defined in Equation 3.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

4. **F1 Score:** A harmonic mean of Precision and Recall, providing a balanced measure of a model's accuracy, especially useful in cases of imbalanced classes. It is calculated as shown in Equation 4.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

In these definitions, TP refers to True Positives, TN to True Negatives, FP to False Positives, and FN to False Negatives.

4 Results

4.1 Overall Performance Overview

All implemented models achieved remarkably high performance, consistently exceeding 97% across all key evaluation metrics. This underscores the effectiveness of the extensive data preprocessing and feature engineering pipeline. As detailed in Table 3, the TCN consistently achieved the highest scores.

Table 3: Overall model performance comparison.

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.9794	0.9797	0.9794	0.9794
ANN	0.9747	0.9750	0.9747	0.9747
TCN	0.9873	0.9880	0.9870	0.9870

4.2 Detailed Analysis of Temporal Convolutional Network (TCN) Performance

The TCN model demonstrated exceptional classification capabilities. Its performance is particularly noteworthy for individual driving styles, as detailed in Table 4. A significant finding is the perfect recall (1.000) for 'AggressiveStyle,' meaning the TCN correctly identified every instance of aggressive driving in the test set. This characteristic is critical for safety-oriented applications. The overall global accuracy of the TCN model was 0.9873.

Table 4: Detailed TCN performance per style.

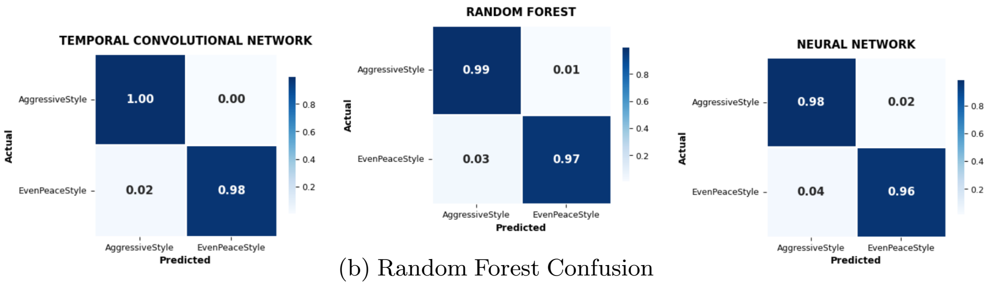
Style	Precision	Recall	F1 Score	Support
Aggressive	0.975	1.000	0.988	119
Even Pace	1.000	0.975	0.987	118
Weighted Avg	0.988	0.987	0.987	237

4.3 Visual Performance Analysis: Confusion Matrices

The classification performance of all models is visually represented through the confusion matrices in Figure 2. The matrices display high true positive rates and minimal misclassifications across all models. As shown, the TCN confusion matrix (Fig. 2a) clearly demonstrates its superior performance.

4.4 Training Dynamics of the TCN Model

The training and validation curves for the TCN model are visualized in Figure 3. These curves illustrate a rapid convergence of accuracy and a consistent decrease in loss, which are strong indicators of efficient learning without significant overfitting. The final metrics, with a training accuracy of 0.986 and a validation accuracy of 0.981, confirm the model's stability and its strong capacity for generalization to unseen data.



(a) TCN Confusion Matrix.

(b) Random Forest Confusion Matrix.

(c) ANN Confusion Matrix.

Fig. 2: Confusion matrices for the (a) TCN, (b) Random Forest, and (c) ANN models. Each matrix compares predicted vs. actual driving styles.

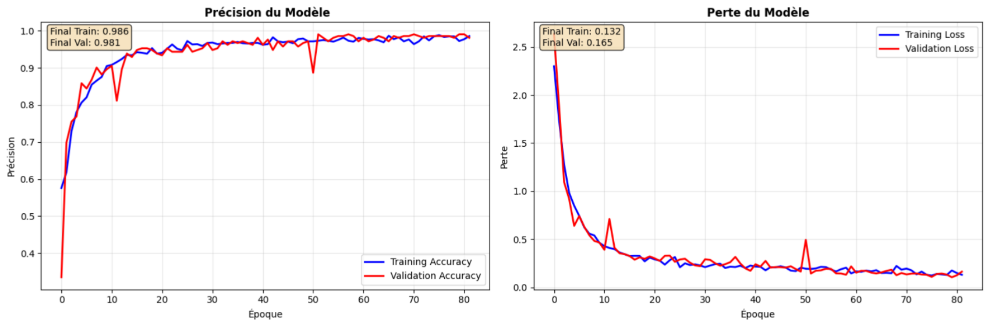


Fig. 3: TCN training curves for accuracy and loss over 80 epochs.

5 Discussion

5.1 Interpretation of Results and Model Strengths

The experimental results consistently highlight the high performance achieved by all models in this study, generally exceeding 97% across key metrics. This strong performance validates the overall effectiveness of the proposed methodology for driving style classification using the Mafalda dataset. Each model contributed unique strengths to this success:

- The Random Forest model proved to be highly robust, efficiently handling non-linear relationships and identifying crucial features within the dataset.
- The Artificial Neural Network demonstrated its ability to learn complex data patterns through its deep representation capabilities.
- The Temporal Convolutional Network (TCN), however, truly excelled, showcasing superior performance in processing sequential data. Its architecture, particularly its dilated causal convolutions, enabled the capture of intricate, long-range temporal dependencies with exceptional precision. This specialized ability for dynamic data analysis positioned the TCN as the most effective model for the classification task.

It is important to reiterate that the high quality of the Mafalda dataset, combined with the meticulous advanced preprocessing pipeline, was fundamental to achieving these impressive results across all models. The careful preparation of data ensured that the models received clean, relevant, and well-structured input.

5.2 Comparison with Prior Work and Demonstrated Superiority

A critical outcome of this study is the significant quantitative improvement over existing benchmarks. Table 5 provides a direct comparison of our models' performance against other relevant findings in the field, contextualizing our contribution. As evident from the table, the models developed in this study, particularly the TCN, achieved a substantial leap in performance. The TCN's F1-score of 0.987 is significantly higher than the 0.85 reported by Al-Refai et al.'s best model [2] and also surpasses other recent deep learning approaches [7, 9, 6].

Table 5: Comparative Performance: This Work vs. State-of-the-Art.

Author(s)	Method	Dataset	F1-Score
This Work	**TCN**	**Mafalda**	**0.987**
Al-refai et al. (2024) [2]	ANN + Bagging	Mafalda	0.85
Khan et al. (2023) [7]	CNN-LSTM	UAH-Drive	0.92
Wang et al. (2022) [9]	LSTM	Private	0.90
Conti et al. (2021) [6]	SVM	Private	0.88

This distinct quantitative advantage confirms the core contribution of the current research. This superiority is attributed to several key factors:

- **Advanced Model Architecture:** The strategic choice and optimized implementation of the TCN enabled the capture of intricate temporal dependencies in driving patterns more effectively than the conventional feed-forward ANNs used in prior work.
- **Enhanced Feature Engineering:** The meticulous pipeline for feature engineering and data preprocessing extracted more discriminative and robust patterns from the raw OBD-II data. This approach extends significantly beyond typical raw data usage.
- **Optimization and Training Regime:** Optimized hyperparameter tuning and a robust training methodology further contributed to the enhanced performance and generalization capabilities of the models.

5.3 Implications of Key Features

Analysis of feature importance yielded a significant finding: kinematic changes, particularly those related to vehicle speed, emerged as more critical indicators of driving style than static engine parameters. This observation suggests that the dynamic profile of a vehicle's movement, including instantaneous speed, speed variation, and average speed, provides more discriminative information for classifying driving behaviors than, for instance, engine RPM or temperature. This insight can profoundly inform the design of future in-vehicle monitoring systems, guiding the selection of relevant sensors and potentially leading to more efficient and targeted data collection strategies.

5.4 Limitations and Future Work

Despite the significant advancements demonstrated, certain limitations offer avenues for future research:

- **Dataset Scope:** The study relied solely on the Mafalda dataset. Future work could explore the generalizability of these models across diverse datasets from different geographic regions or vehicle types.
- **Real-time Deployment:** While the models show high accuracy, evaluating their performance and computational efficiency in real-time embedded systems remains a crucial next step.
- **Explainable AI (XAI):** Exploration of Explainable AI (XAI) techniques is vital for practical applications, particularly in safety-critical domains where understanding model decisions is paramount.
- **Multi-modal Data Fusion:** Integrating data from additional sensors (e.g., GPS, accelerometer from smartphones, or in-cabin cameras) could provide richer contextual information and potentially lead to even more nuanced driving style classification.

6 Conclusion

This study successfully presented a comprehensive approach to advanced driving style classification using OBD-II data, employing Temporal Convolutional Networks (TCNs), Random Forest (RF), and Artificial Neural Networks (ANNs). The enhanced data preprocessing and sophisticated feature engineering pipeline, coupled with the power of TCNs, resulted in consistently high performance metrics, all exceeding 97%. Notably, the TCN demonstrated superior capabilities in capturing complex temporal dependencies, achieving a perfect recall for aggressive driving, a critical aspect for safety applications. The significant quantitative improvements over existing benchmarks underscore the efficacy of the proposed methodology and the potential of TCNs in this domain. These findings contribute substantially to the field of Intelligent Transport Systems, paving the way for more accurate driver behavior monitoring and personalized vehicle intelligence.

References

[1] Michele Ruta, Floriano Scioscia, Giuseppe Loseto, Agnese Pinto, and Eugenio Di Sciascio. Machine learning in the internet of things: A semantic-enhanced approach. *Semantic Web*, 10(2):183–204, 2019.

[2] Ghaith Al-refai, Mohammad Al-refai, and Arafat Alzu'bi. Driving style and traffic prediction with artificial neural networks using on-board diagnostics and smartphone sensors. *Applied Sciences*, 14(12):5008, 2024.

[3] Paul Baltusis. On board vehicle diagnostics. In *SAE Technical Paper*. SAE International, 2004.

[4] S. Suthaharan. Support vector machine. In *Machine Learning Models and Algorithms for Big Data Classification*, pages 207–235. Springer, New York, NY, USA, 2016.

[5] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[6] M. Conti and S. Rossi. A support vector machine approach for aggressive driving pattern recognition. *Transportation Research Part C: Emerging Technologies*, 128:103123, 2021.

[7] A. Khan and S. Lee. Hybrid cnn-lstm models for real-time driver behavior classification. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):1121–1135, 2023.

[8] J. Pérez, C. Fernández, and R. Pérez. The uah-drive dataset: A new comprehensive dataset for the study of naturalistic driving. In *Procedia Computer Science*, volume 52, pages 517–523. Elsevier, 2015.

[9] L. Wang and Y. Chen. Driving style analysis using long short-term memory networks on telematics data. In *Proceedings of the International Conference on Machine Learning and Cybernetics*, pages 345–352, 2022.

[10] Satyam Sharma, Shunnu Sharma, and Anish Athaiya. Activation functions in neural networks. *Towards Data Science*, 6:310–316, 2017.

[11] Simon Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, 1994.

[12] Sebastian Bock and Martin Weiß. A proof of local convergence for the adam optimizer. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[13] V Roshan Joseph and Akhil Vakayil. Split: An optimal method for data splitting. *Technometrics*, 64(2):166–176, 2022.

[14] David MW Powers. Evaluation: From precision, recall and f-measure to roc, informedness, markedness correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.