International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Traffic Anomalies Detection Using Machine Learning and Deep learning

Surya Chittimoju

Student, Rajam, Vizianagaram, 532127, India.

ABSTRACT

Data science is a field of study that works with enormous amounts of data using cutting-edge tools and methods to uncover hidden patterns, gather useful information, and make business decisions. The advancement of 5G has made it possible for Autonomous Vehicles (AVs) to have total control over every aspect of the operation. To be able to move independently, the AV takes autonomous actions and gathers travel information using a variety of smart devices and sensors. A computational data science strategy (CDS) is suggested for managing vast amounts of traffic data in various formats. The computational data science method was developed to find anomalies in traffic data that impair traffic efficiency. The integration of data science and cutting-edge AI methods, including deep learning, leads to a better degree of data anomaly identification, which reduces traffic jams and vehicle queues. The early identification of the variables that led to data abnormalities to prevent long-term traffic jams may be summed up as the primary contribution of the CDS technique. Additionally, CDS showed encouraging outcomes in many scenarios involving road traffic.

Keywords: Computational Data Science, Deep Learning, Autonomous Vehicle, AI, Data Anomalies.

1. Introduction

The evolution of 5G networks is poised to revolutionize autonomous vehicle (AV) development, unlocking new possibilities. A critical step involves verifying diverse data types for accuracy, integrity, availability, and inconsistencies. Geographical factors affecting datasets must be considered, especially since incomplete or falsified road data reaching an AV can lead to abnormal conditions in urban settings. To tackle this, employing a deep learning model for meticulous dataset control is recommended.

Deep learning, a versatile tool across various research domains, requires substantial datasets for effective model development. This study concentrates on identifying and classifying anomalies in road traffic, steering away from the laborious and costly traditional methods of surveying road conditions. Collaborative mobile sensing, leveraging data collected by smartphones and employing data-mining approaches, emerges as a promising alternative for detecting and classifying road anomalies. The data science life cycle guides the analysis, encompassing phases like data collection, preparation, exploration, modeling, evaluation, and deployment. Anomaly detection methods fall into two main categories: model-based and data analysis-based. While data analysis relies on statistical measurements, model-based approaches utilize precise algorithms like machine learning schemes. Urban road anomalies, causing discomfort to drivers and disrupting traffic flow, often stem from accidents, congestion, unfavorable weather, construction projects, and frequent lane changes within AV networks. The introduction of visual surveillance to identify data anomalies in road traffic marks a notable advancement. Yet, challenges persist, including effective communication between AVs in heterogeneous wireless networks. Heterogeneity introduces delays, necessitating middleware for seamless integration across diverse networks offering varying quality of services (QoS) for AV communication. The identification of road segments with excessive noise poses an additional hurdle for AVs. Various forecasting techniques and recent deep-learning approaches aim to predict urban traffic flow, contributing to anomaly detection. This work adopts a computational data science approach, implementing deep learning schemes to pinpoint interference and other sources of abnormalities in road traffic.

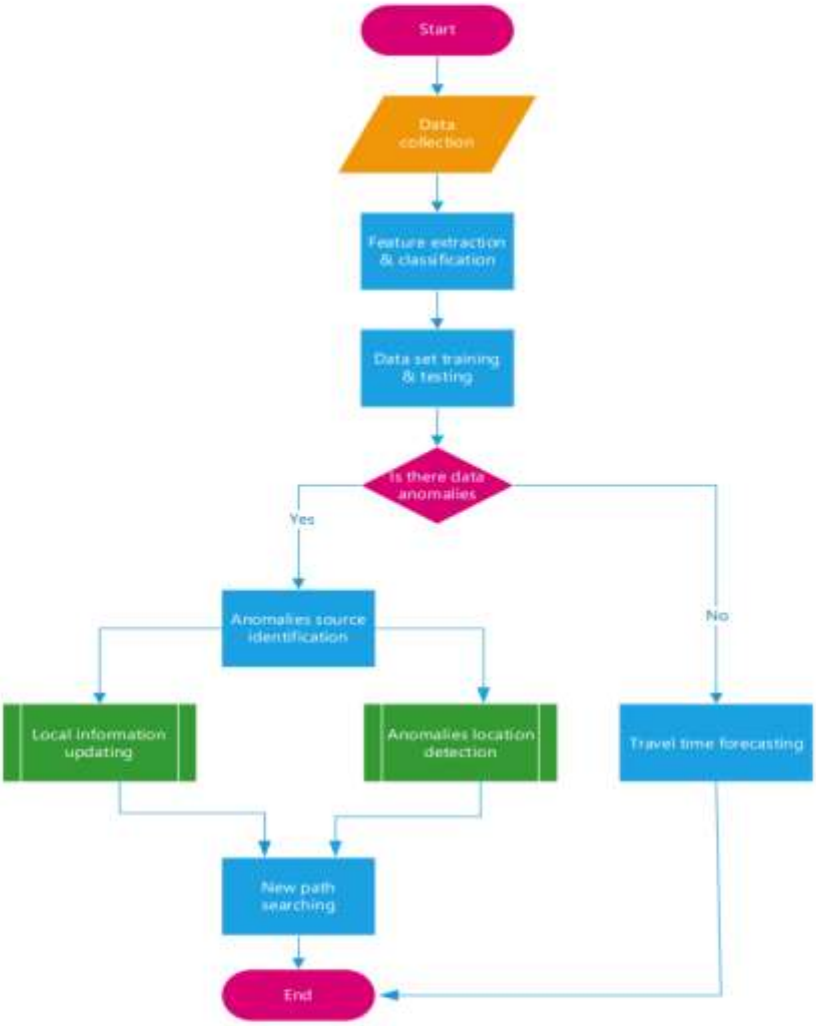
Start
Data collection
Feature extraction & classification
Data set training & testing
Is there data anomalies
Yes
No
Anomalies source identification
Local information updating
Anomalies location detection
Travel time forecasting
New path searching
End

Fig-1: - CDS Algorithm

2. Literature Survey

In his groundbreaking work [1], Jamal Raiyn explores the transformative impact of the 5G era on autonomous vehicles (AVs), enabling them to autonomously gather data and perform sophisticated data analysis using state-of-the-art AI technology. The methodology employed is a Computational Data Science (CDS) approach, incorporating the data science lifecycle's key stages: Data collection, preparation, availability, exploration, modeling, and evaluation, culminating in deployment.

This study introduces a Deep Learning (DL) method specifically designed for the early detection of traffic anomalies, involving a comprehensive process encompassing dataset preparation, training, testing, and performance metrics evaluation. The proposed DL model is expressed by the formula yAD = åxnwn + error, where xn represents input signals, wn signifies weights corresponding to each input signal, and yAD denotes the output signal.

To delve into the intricacies of travel data based on mobile services, the study employs linear regression. The innovative system presented in this paper aims to identify anomalies in traffic flow that contribute to congestion. Through the application of DL principles, this system achieves early detection capabilities for both traffic congestion and accidents, showcasing a forward-looking approach to addressing critical challenges in the realm of autonomous vehicles and intelligent transportation systems..

In paper [2] Y. Li, T. Guo, R. Xia and W. Xie introduced a theory by considering a large amount of uncertain information existing in traffic surveillance videos an algorithm was proposed for traffic anomaly detection for straight roads based on fuzzy theory. The traffic anomaly detection algorithm contributes:

- The fuzzy traffic flow
- The fuzzy traffic density

- The target's fuzzy motion state
- The traffic anomaly detection algorithm was proposed

Finally, experiments show that the algorithm can accurately detect traffic anomalies. It has a high accuracy rate and strong robustness.

In their notable contribution [3], Y. Yao and team introduced an innovative unsupervised approach for detecting anomalies in traffic videos (VAD) by focusing on future object localization. They went a step further by employing an ensemble method that cleverly combines both object- and frame-level VAD techniques to enhance overall performance. The paper introduces the DoTA dataset, which not only includes temporal, spatial, and categorical annotations but also proposes a novel spatial-temporal Area Under the Curve (STAUC) metric for a more nuanced evaluation of VAD effectiveness. Beyond anomaly detection, the DoTA dataset proves versatile, enabling research in video action recognition (VAR) and online action detection within driving scenarios. The authors also hint at future work involving the early detection of traffic accidents and the validation of autonomous driving systems.

In another significant contribution [4], K. K. Santhosh, D. P. Dogra, P. P. Roy, and A. Mitra propose a unique color gradient representation for vehicular trajectories extracted from static camera-recorded videos. The trajectories are meticulously classified, allowing for the detection of anomalies such as lane violations, sudden speed changes, abrupt terminations, and vehicles moving in the wrong direction. The authors introduce a hybrid CNN-VAE (Convolutional Neural Network - Variational Autoencoder) detector, leveraging the strengths of both architectures. A semi-supervised labeling technique, employing a modified Dirichlet Process Mixture Model (mDPMM) clustering, is presented for preparing training data. The hybrid architecture not only detects anomalies but also classifies them, addressing a critical limitation of standalone VAE models.

In the insightful work presented in [5], Tišljarić, Leo, Sofia Fernandes, Tonči Carić, and João Gama propose a tensor-based method for extracting spatiotemporal road traffic patterns to identify anomalies in urban road networks. Their focus lies on two specific anomalies: bottleneck starts characterized by sudden braking in transition and intense acceleration in transition when leaving congested regions. The method involves computing Spatiotemporal Matrices (STMs) and utilizing a tensor composed of these STMs to model the intricate traffic patterns, considering the inherent spatiotemporal nature of traffic data. The authors evaluate the anomaly detection results on segments of the urban road network in a medium-sized European city, providing valuable insights into the detection of nuanced traffic anomalies..

3. Data Collection

Autonomous vehicles (AVs) have at their disposal a range of tools for collecting data, setting them apart from traditional methods like magnetic sensors and on-site surveyors commonly used in traffic data gathering. AVs leverage state-of-the-art technology to overcome the limitations associated with these conventional approaches. This paper delves into a method for data collection that relies on mobile services. In this process, AVs receive data from different devices in diverse formats, introducing challenges related to data quality.

During the preparation phase, a careful procedure is implemented to ensure data accuracy. This involves converting the data into the desired format, followed by a thorough cleaning of the dataset. In this cleanup stage, efforts are made to eliminate any inconsistent, invalid, or corrupt data, ensuring the reliability of the dataset. It's noteworthy that a significant portion of the datasets utilized in this study is openly available to the public, underscoring. transparency and facilitating reproducibility in research efforts.

4. Computational Data Science Approach (CDS): -

In this study, we adopted a computational data science approach to identify anomalies in road travel data. This method blends the principles of the data science lifecycle with advancements in artificial intelligence methodologies.

4.1 Data Science Lifecycle:

The data science lifecycle encompasses key analytical steps based on artificial intelligence methodologies. These steps include data collection, data preparation, data exploration, and data validation.

- Data Collection:

Autonomous vehicles (AVs) leverage a diverse array of tools for data collection, diverging from traditional methods involving human surveyors and magnetic sensors. This paper introduces a novel approach to data collection centered around mobile services.

- Data Preparation:

AVs receive data in various formats from different devices, posing challenges to data quality. In the preparation phase, efforts are made to convert the data to a standardized format. Subsequently, the dataset undergoes a thorough cleaning process, eliminating inconsistent, invalid, and corrupt data.

- Data Availability:

The acquired trip dataset contains essential statistical metrics.

- Data Exploration:

Our exploration of the data focuses on understanding the dynamics of traffic flow on urban roads. Visualization tools, such as plotting, reveal hidden patterns within the dataset. Additionally, various statistical measures, including mean and standard deviation, along with their interactions with other features, aid in distinguishing between normal and abnormal traffic flow. To identify abnormal situations, we observe a decrease in traffic load (tt(t, k)), and as issues resolve, the traffic load progressively increases. An abnormal record is defined when the traffic speed drops at least 30 km/h below the average speed for the same day of the week and time. This 30 km/h threshold symbolizes the minimum speed change considered "abnormal," and the determination of this threshold relies on observed travel data.

5. Computational Artificial Intelligence:

In the realm of defining agents, numerous perspectives exist, often rooted in the functions and behaviors exhibited within specific domains. Essentially, an agent's operational definition hinges on the environment where it delivers its services. Key characteristics attributed to an agent include autonomy, indicating its ability to initiate actions and wield control, managing negotiations with humans or other agents to enhance rules. Additionally, agents employ reasoning strategies to make informed decisions.

Broadly speaking, artificial intelligence (AI) denotes a computer or machine's capacity to emulate human cognitive abilities. Establishing a relationship among the terms artificial intelligence (AI), machine learning (ML), and deep learning (DL) unfolds as follows: AI encompasses diverse fields involving expert systems that make decisions based on intricate rules. Within this landscape, ML surfaces as a subset of AI, where systems autonomously learn and adapt, progressively refining their performance on specific tasks as they assimilate more data. On the other hand, deep learning technologies, another subset of machine learning applications, signify a realm where systems teach themselves to excel at particular tasks with increasing precision, all without human intervention.

5.2 Deep Learning Technology

Deep Learning (DL) stands out as a powerful tool in the realm of machine learning, revolutionizing how we approach complex problem-solving. Its methodology involves processing data in a hierarchical manner, gradually incorporating more abstract and invariant properties at each level. DL excels at learning the intricate features of a dataset and cleverly combining them to achieve specific goals. In the context of addressing challenges within intelligent transportation systems, the DL approach emerges as a key player. This research advocates for employing a DL approach to proactively identify traffic irregularities. The process involves two fundamental components: training and testing. The proposed system comprises four distinct phases: (1) preparing the dataset, (2) initiating a training phase, (3) conducting a testing phase, and (4) evaluating performance metrics. Travel speed data, crucial for this study, is sourced through smartphone services. The data undergoes essential pre-processing phases, including cleaning and recovering missing values.As the DL model gears up in the training phase, it delves into the features extracted from the pre-processing steps. This dynamic learning process equips the system to navigate and comprehend the complexities of the dataset, paving the way for effective problem-solving and early identification of traffic irregularities..

5.3 Architecture of DL Concept

An input layer, a hidden layer, and an output layer are the three major components of the DL in general.

• The Input layer

The input layer for DL comprises a huge amount of data that is acquired from numerous sources. The big dataset for traffic modelling is diverse and comes from a variety of sources, such as cameras, LIDAR, sensors, and GNSS. Historical traffic statistics and near-real-time data are provided by the equipment placed in AVs.

• The Hidden layer

The hidden layer is responsible for processing the input data. It processes the attributes of the dataset and extracts useful information to construct new attributes that will be used as input for the DL model. Each layer within the hidden layer is assigned rules focused on input data attributes, which are updated in keeping with new data input. The size of the hidden layer is expressed in terms of the number of neurons there. The neurons have an important influence on the learning ability of the algorithm; too few can lead to insufficient learning, and too many can lead to overfitting.

• The Output layer

The output layer is responsible for exporting the values, or the vectors of the values, that correspond to the format required for the problem, and it presents the visual results based on measurements of statistical error.

5.4 Urban Road Anomalies

Anomaly detection revolves around identifying data that deviates from the norm. In the context of urban road travel, the significance lies in promptly addressing these outliers. The proposed approach centers on the notion that monitoring changes in the behaviour of individual autonomous vehicles (AVs), such as deceleration and lane changing, can unveil traffic anomalies. The anomaly detection scheme unfolds in key stages, commencing with the feature extraction stage. This step involves transforming original traffic variables into features, encapsulating crucial information for the detection task. In this study, the feature extraction leverages a deep learning scheme. The architecture comprises three layers: an input layer, a hidden layer, and an output layer. The initial phase, the training phase, draws insights from the raw travel dataset, incorporating attributes like time, road section, and speed. Speed observations, gathered through mobile services at 2.5-minute intervals, lay the foundation for the model. The deep learning architecture unfolds with an optimal selection of neurons in the hidden layer, mirroring the number of urban road sections, each spanning 300 meters. The subsequent layer in the hidden stage features a reduced number of neurons, contributing to the intricate learning process of the model.

6. Methodology

Navigating through the anomaly detection process presents a myriad of challenges, primarily revolving around the task of pinpointing patterns in the data that deviate from the expected norm. The process kicks off with establishing a baseline for what constitutes typical traffic behavior on a stretch of urban road. Observations that fall outside this customary pattern are then singled out as anomalies. The real hurdle lies in unraveling these distinctive patterns, forming the crux of anomaly identification in city road traffic. Notably, the computational data science (CDS) methods employed in this study demonstrate their versatility, applicable to a wide array of regularly traversed roadways, setting them apart from earlier approaches.

6.2 Description of DCS

In the quest to identify anomalies in road traffic data, numerous algorithms have been crafted. These algorithms primarily target the detection of data anomalies that materialize as traffic congestion. In contrast, the operations of the computational data science (CDS) algorithm are centered around understanding the intricate behaviours exhibited by data, influenced by a multitude of factors. Cognitive data goes a step further by discerning between the impact of external elements like cyber attackers, geographical conditions, and radio channel interference, as well as internal factors such as delays in vehicle-to-vehicle communication triggered by changes in Quality of Service (QoS) requirements. While traditional algorithms, like Machine Learning (ML), often focus on describing the structured traffic data at hand, CDS algorithms take on the role of sensing the root causes of anomalies and pinpointing their locations.

6.3 Data Anomaly Detection Based on CDS

Assessing travel data and Autonomous Vehicle (AV) location information involved employing various methods. The pivotal challenge in this endeavor was finding the right input dataset, leading to an extensive preparation process for the massive data influx:

- **Data Recognition:**

AVs are equipped with diverse devices for collecting data. To discern between different types of datasets, classification tools were essential. The study utilized smartphones and Ublox devices to capture vehicle positioning information.

- **Structured vs. Unstructured Data:**

AVs typically gather organized data, often represented in matrices where each column signifies a different property. For instance, the positioning data set included columns for longitude and additional properties. Addressing unstructured data from certain AV devices involved the initial step of organizing it into a matrix.

- **Cleaning and Formatting:**

Deduplication and formatting played a crucial role. Various devices collecting travel data are influenced by environmental and anthropogenic factors. Data cleaning involved editing to remove irrelevant or inaccurate travel observations. Optimal computational data formats were adopted to enhance processing ease, readability, and compatibility with different tools and systems. The travel data, sourced from various smart devices in different formats, underwent cleaning to enhance analysis effectiveness and result quality.

- **Visualizing Data:**

An interactive exploratory journey dataset proved valuable for real-time fault visualization. Some gathered datasets were incomplete due to noise and environmental conditions affecting data transfer. Identifying fields with missing data and appropriately compensating for them became a crucial aspect of data cleansing.

- **Ranking of Data:**

Historical data underwent grading and ranking using tools designed to identify missing data and errors. Performance metrics, evaluating availability, accuracy, and integrity, were employed to grade road segments, providing an overall measure of different devices' effectiveness.

6.4 Anomaly detection based on DL

The properties were extracted from the input data by the deep learning algorithm. The characteristics were categorised, and scores were given.

$y_{AD} = \sum x_n w_n + \text{error}$

Here, x_n shows the input signal, wn shows the weight corresponding to each input signal, and y_{AD} shows the output signal. FDL is the activation function which means calculating the sum of data coming from the input.

6.5 Anomaly Detection Schemes Evaluation

Autonomous Vehicles (AVs) routinely collect diverse traffic data, primarily sourced in 2.5-minute intervals from smartphones. However, upon statistically analyzing the raw traffic data, two significant issues emerged: the dataset was incomplete, and noise was present in the data.

In urban settings, roads are conveniently divided into segments, each spanning 300 meters. To refine the data and enhance its reliability, various processes were applied, including the removal of extraneous components such as noise.

The importance of high-quality data cannot be overstated, as low-quality data may contribute to accidents and traffic congestion. Additionally, urban noise, particularly in network tunnels, may obstruct the reception of complete data. To discern and address these issues, statistical metrics were employed to identify incomplete data and refine the overall data quality..

7. Results and Discussion

For data analysis, linear regression is frequently employed. The travel data based on mobile services were analysed in this study using linear regression. Furthermore, it is advised to estimate and measure the impact of noise on travel data that is gathered by mobile services using linear regression and Pearson linear correlation. An analysis of statistics is shown in Table 1.

REGRESSION STATISTICS

	Without noise	noise
MULTIPLE R	0.323592	0.940014072
R SQUARE	0.104908	0.902426765
ADJUSTED R SQUARE	0.100455	0.901694658
STDEV	0.032098	10.25988381
SKEWNESS	0.610393861	0.934475831

Table 1: Regression Statistics

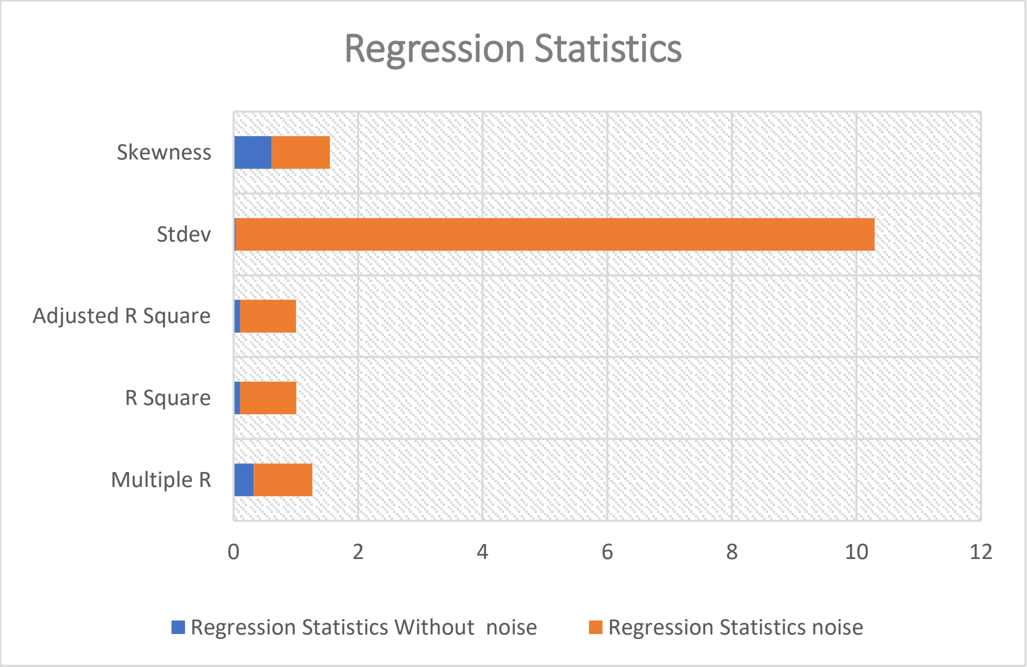

Fig 2: Graphical Representation of Regression Statistics

In the realm of data analysis, linear regression stands as a commonly employed technique. In this study, an analysis of travel data sourced from mobile services was conducted utilizing linear regression. Moreover, it is recommended to employ linear regression and Pearson linear correlation to gauge and assess the influence of noise on the travel data collected through mobile services.

The findings of our study on the computational data science approach reveal a promising avenue for addressing challenges in various domains. By integrating the principles of the data science lifecycle with advancements in artificial intelligence, our approach offers a nuanced understanding of anomalies in road travel data. The application of deep learning models in anomaly detection showcases its efficacy, particularly in identifying irregularities in traffic flow. One of the key contributions lies in the successful utilization of mobile sensing and collaborative data mining to automatically classify road anomalies. This not only streamlines the traditionally time-consuming and expensive road condition survey methods but also enhances the accuracy of anomaly detection. The proposed model, combining artificial intelligence and goal-oriented agents, demonstrates adaptability and autonomy in handling diverse datasets affected by geographical factors. Our study delves into the intricacies of the data science lifecycle, emphasizing the importance of data collection, preparation, exploration, modelling, and evaluation. The utilization of deep learning schemes proves instrumental in training effective models for anomaly detection, especially in the context of urban road traffic. Furthermore, the research introduces a comprehensive methodology for addressing challenges related to communication between autonomous vehicles in heterogeneous wireless networks. Middleware adoption is proposed to ensure seamless communication in networks with varying quality of services (QoS) for autonomous vehicle communication.

In conclusion, our computational data science approach contributes to the advancement of anomaly detection in road traffic, paving the way for more efficient and reliable autonomous vehicle systems. The combination of deep learning, collaborative mobile sensing, and goal-oriented agents represents a holistic and effective strategy for tackling anomalies in urban road networks.

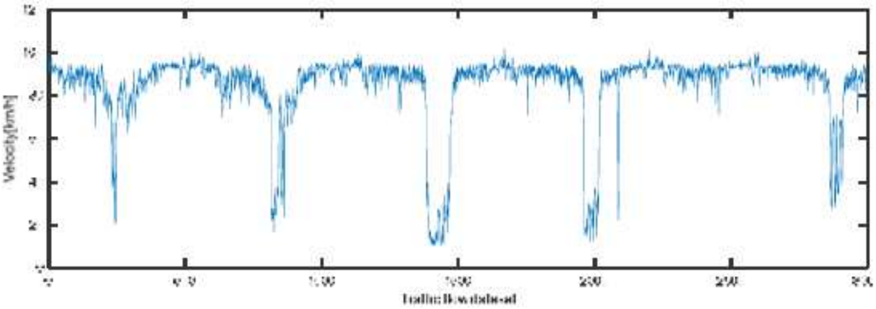

Fig 1: Positioning Data Anomaly Detection Based on DL

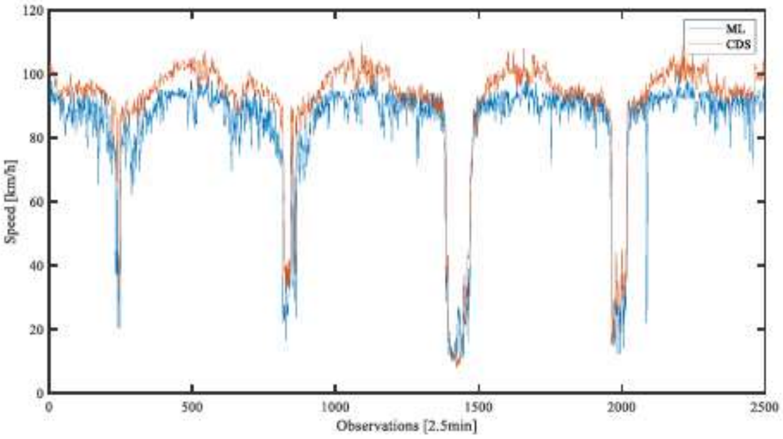

Fig 2: Comparison between CDS and ML

8. Conclusion

This study gives a solution to overcome the problems faced by Autonomous vehicles by pre-processing the original raw data. Missing information was compensated for with a data-cleaning process. This reduced or eliminated unwanted features attributed to noise in the original data. The processed dataset was divided into training and testing subsets to carry out supervised learning. In this paper, a novel system is proposed to detect anomalies in traffic flow that lead to congestion. The DL concept, based on statistical measurements, makes possible the early detection of traffic congestion and traffic accidents.

Thus, the proposed system may have a direct and significant positive impact on drivers' health and safety. The simulation results demonstrate that the forecasting system was improved by the use of the DL network.

REFERENCES

[1]. Jamal Raiyn. Detection of Road Traffic Anomalies Based on Computational Data Science, 11 January 2022, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1149975/v1]

[2]. Y. Li, T. Guo, R. Xia and W. Xie, "Road Traffic Anomaly Detection Based on Fuzzy Theory," in IEEE Access, vol. 6, pp. 40281-40288, 2018, doi: 10.1109/ACCESS.2018.2851747.

[3]. Y. Yao et al., "DoTA: Unsupervised Detection of Traffic Anomaly in Driving Videos," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2022.3150763.

[4]. K. K. Santhosh, D. P. Dogra, P. P. Roy and A. Mitra, "Vehicular Trajectory Classification and Traffic Anomaly Detection in Videos Using a Hybrid CNN-VAE Architecture," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2021.3108504.

[5]. Tišljarić, Leo, Sofia Fernandes, Tonči Carić, and João Gama. 2021. "Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach" Applied Sciences 11, no. 24: 12017, doi:10.3390/app112412017.

[6]. Aditya Vikram, Mohana, "Anomaly detection in Network Traffic Using Unsupervised Machine Learning Approach," available at researchdate.net June 2020, doi: 10.1109/ICCES48766.2020.9137987.

[7]. Feng, R.; Yao, Y.; Atkins, E. Smart Black Box 2.0: Efficient High-Bandwidth Driving Data Collection Based on Video Anomalies. Algorithms 2021, 14, 57. https:// doi.org/10.3390/a14020057.

[8]. Qingchen Zhang, Laurence T. Yang, Zhikui Chen, Peng Li, "A survey on deep learning for big data," available at 11 November 2017 1566-2535/2017 Elsevier, doi: http://dx.doi.org/10.1016/j.inffus.2017.10.006.

[9]. Fotiadou, K.; Velivassaki, T.-H.; Voulkidis, A.; Skias, D.; Tsekeridou, S.; Zahariadis, T. pfSense Network Traffic Anomaly Detection via Deep Learning. Information 2021, 12, 215. https://doi.org/10.3390/ info12050215