Journal of the Society of Automotive Engineers Malaysia
Volume 10, Issue 1, pp 32-39, January 2026
e-ISSN 2550-2239

Unveiling Motorcyclist Behaviour Using Naturalistic Riding Study

H. L. Khoo*[1] and C. H. Teoh[2]

[1] Centre for Sustainable Mobility Technologies, Universiti Tunku Abdul Rahman (UTAR), Sg. Long Campus, 43000 Kajang, Selangor, Malaysia
[2] School of Computer Science, Faculty of Innovation & Technology, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia

*Corresponding author: khoohl@utar.edu.my

ORIGINAL ARTICLE *Open Access*

Article History:

Received
28 Oct 2025

Accepted
15 Dec 2025

Available online
1 Jan 2026

ABSTRACT – *Motorcyclists represent a significant portion of road users in Malaysia, where motorcycle-related accidents contribute disproportionately to road fatalities and injuries. Traditional data collection methods, such as crash reports and simulations, often lack the contextual detail and behavioural nuance necessary to understand rider risk exposure. This study employs the Naturalistic Riding Study (NRS) approach to collect and analyze the riding behaviour of motorcyclists in urban and suburban areas in the Klang Valley region, Malaysia. To facilitate data collection, an Internet-of-Things (IoT) system is developed to detect real-time riders' behaviour. This paper illustrates the development of the system and presents some data collected during the pilot test.*

KEYWORDS: Naturalistic riding behaviour, motorcycle, risk, accident, safety, IoT

Journal homepage: www.jsaem.my

1. INTRODUCTION

Motorcycle accidents pose a severe road safety issue in Malaysia, accounting for a disproportionate share of road fatalities and injuries. Statistics show that motorcyclists account for over 60% of road traffic fatalities in Malaysia annually (MIROS, 2022). More recent tallies show the share rising to around 70–75% in 2023–2024 (≈4,489 of 6,473 deaths in 2024), underscoring a worsening exposure-risk profile for motorbikes (MOT, 2022; Jarod, 2024). Factors contributing to these accidents include risky riding behaviors such as speeding, abrupt lane changes, and failure to observe traffic rules, compounded by environmental factors like road conditions, weather, and traffic congestion (Abdul Manan & Várhelyi, 2012). More recent Malaysian studies identify behavioural and system factors: excessive speed and red-light running; risky merges and lane-filtering; young riders' over-representation; primary-road exposure; intersection conflicts; and time-of-day patterns. These are consistently observed in national crash analyses and observational work on access points and primary roads (Manan, Khaidir, & Jamil, 2020). The complexity of motorcycle crashes requires in-depth studies to investigate the interaction between the rider's behaviour and accident causation.

When dealing with rider behaviour analysis, conventional methods, such as crash data analysis, simulations, interviews, and questionnaires, are limited in their ability to provide a full picture of accident risk. The naturalistic riding studies (NRS) method, on the other hand, offers an opportunity to understand the near-real behaviour of riders. The NRS utilizes instrumented vehicles to collect rich and context-sensitive data. In some cases, near-miss data could be collected as well. Some past studies utilize the NRS method to collect rider behavioural data, e.g., Williams (2014) conducted the MSF-100 Motorcyclists Naturalistic Study in the USA, Halim et al. (2025) in Indonesia, Cho et al. (2023) and Kim et al. (2024) in South Korea, and Espié (2013). These studies gathered detailed behavioral data through sensors and cameras to capture the rider's essential behaviors such as speed, braking, acceleration, and riding posture in urban traffic. This allows the researchers to perform behavioural analysis of the rider's interactions with road infrastructure and traffic conditions.

Recognizing the need for detailed rider behavior data in Malaysia, this study adopts the naturalistic riding study (NRS) methodology to record the riders' behaviour in the urban and suburban settings within the Klang Valley region. The NRS employs an Internet-of-Things (IoT) system specifically developed to monitor and record real-time motorcycle rider behaviour under different road geometries and traffic conditions. Our IoT-based platform provides participant motorcycles with synchronized inertia measurement unit (IMU), i.e. roll/pitch/yaw, lateral/longitudinal gravity; and Global Navigation Satellite System (GNSS) or Global Positioning System (GPS); the data are timestamped and streamed/logged through local caching and Message Queuing Telegraphic Transfer (MQTT)/Narrowband IOT (NB-IOT) or Long-Term Evolution (LTE)) for later annotation. The primary objectives of this study are to develop and validate this IoT-based monitoring system and to demonstrate its applicability using preliminary pilot data. This paper provides a comprehensive overview of the system development process that shows the potential of NRS methodologies to enhance motorcycle safety initiatives.

2. METHODOLOGY

This section illustrates the methodology adopted to develop the IOT system for the NRS study.

2.1 Sensor

An IOT system comprising various sensors is integrated to capture a rider's state (3D position, 3D orientation/angles, linear/angular velocities, and others) during naturalistic riding. The sensors include GPS-GNSS receivers, which are used to capture the rider's 3D position (latitude, longitude, and altitude). The GNSS can also calculate the relative ground speed and heading of the rider. The downside is that the GPS requires an unobstructed line of sight. Signal multipath may appear in urban canyons, thus hampering the usability of the GPS-GNSS system.

Inertial Measurement Unit (IMU): An IMU measures angular velocity with respect to 3D planes. IMU consists of 3-axis accelerometers and gyroscopes. The MPU-6050 6-DOF IMU, with an accelerometer integrated with a gyroscope, is used in this study. The technical specifications of the sensors above are detailed in the subsections below.

2.1.1 GPS-GNSS Specifications

The specifications of the SIM A7670X LTE + GPS/GNSS Transceiver are shown in Table 1.

TABLE 1: Technical specifications of the SIM A7670X transceiver

Feature		Description
Radio Frequency Bands	GSM	GSM850, EGSM900, DCS1800, PCS1900
	LTE-FDD	B1, B2, B3, B4, B5, B7, B8, B28, B66 (suitable for Malaysia)
	Category	CAT1
	GNSS	GPS – 1575.42±1.023MHz GLONASS – 1597.5~1605.8MHz BEIDOU – 1561.098±2.046MHz Hot Start TTFF < 1s Cold start TTFF < 40s Accuracy < 2m
Antenna		1. GSM/LTE and 2. GNSS L1 antenna
Interfaces	USIM	1.8V / 3V
	USB	USB 2.0 compliant
	UART	Baud rate from 300bps to 686,400bps, RTS/CTS flow control
Application Protocols		AT (including custom LBS command) GNSS – NMEA HTTP (S), TCP/IP MQTT(S)

2.1.2 IMU Specifications

The specification of the InvenSense's MPU-6050 IMU is shown in Table 2.

TABLE 2: Technical Specifications of the InvenSense's MPU-6050 IMU

Features	Description
Chip Model	MPU-6050
Power	3V – 5V
Communication Protocol	I^2C
Accelerometer Range	+/- 2, +/- 4, +/-8, +/-16 g
Gyroscope Range	+/-250, +/-500, +/-1000, +/-2000 °/s
Resolution	16-bit (AD Converter and data output)

2.1.3 Message Specifications

The data format of the transmitted MQTT Message is shown in Table 3.

TABLE 3: Data Format of the MQTT Message

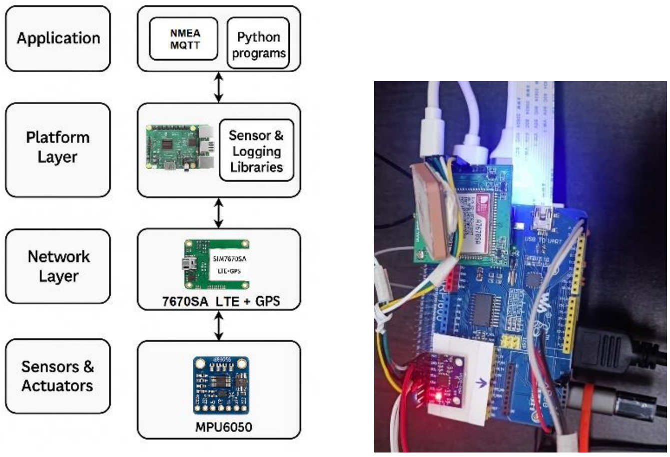
No	Data attributes	Label	Remarks / UOM
1	Date & Time	Datetime	
2	Latitude	Lat	° N/S
3	Longitude	Lon	° E/W
4	Altitude	Alt	M
5	Speed	Speed	km/h
6	Course	Course	0 – 359°
7	Satellite status	CGNSSINFO	* NMEA to receiver-specific format
8	Vehicle	vehicle	
9	Temperature	temperature	°C
10	Yaw rate	yaw	rad /s derived from computation
11	Pitch rate	pitch	rad /s derived from computation
12	Roll rate	roll	rad /s derived from computation
13	Longitudinal acceleration	Ax	g can be converted to ms^{-2} by multiplying 9.81
14	Lateral acceleration	Ay	g
15	Vertical acceleration	Az	g
16	X-axis Angular rate	Gx	° /s
17	Y-axis Angular rate	Gy	° /s
18	Z-axis Angular rate	Gz	° /s

2.2 System Architecture

The key components used in this study include a Raspberry Pi 3B+ single board computer (SBC) running on Raspberry Pi OS (Raspberry Pi Ltd, n.d.) – either Bullseye or Bookworm, a data communication transceiver, a positioning module, with optional CSI camera module (Arducam IMX219 or IMX519) for recording driving behaviors, 10,000mAh power bank, and a USB 3.0 thumb drive for data storage. For data communications, two viable technologies are shortlisted: NB-IOT and LTE. Although NB-IOT uses very little power during communication, it lacks mobility support and has a very poor handoff mechanism compared with LTE. From physical experiments using SIM7000E (NB-IOT) and A7670SA (LTE) communication modules in the Klang Valley, it is evident that LTE has a larger coverage area than NB-IOT. Hence, the LTE module serves as the communication link in the IoT system. The satellite positioning modules usually support GPS/GNSS. Some can support GLONASS/GALILEO and BEIDOU. The module that offers positioning capability together with data communication, i.e., SIM7670SA (LTE + GPS), is used in this study.

The IOT system architecture developed for the NRS study is shown in Figure 1 (a). It consists of 4 layers: the sensor and actuator layer, the network layer, the platform layer, and the application layer. The Sensor and Actuator layer involves acquiring real-time data directly from the physical environment. Devices selected for this study include the *GY-521 MPU6050 6DOF Accelerometer + Gyro,* which provides essential motion-sensing data, including acceleration, angular velocity, and orientation, critical for identifying rider maneuvers and behaviors. The *Neo-M6 GPS + microSD Shield* for Arduino was selected to facilitate precise geolocation data logging and local data storage capability, ensuring continuous location tracking. The Raspberry Pi 3 Model B+ is employed in this study as the primary computational unit, integrating sensor data, executing preliminary data processing tasks, and managing data transmission.

The Network Layer ensures seamless data transmission from the sensor devices to the cloud. The *NB-IOT/GPS HAT SIM7000E for Raspberry Pi* is used in this study to provide reliable cellular connectivity optimized for IoT applications. The advantage of the device is that it enables efficient, low-power communication, making it suitable for transmitting sensor data to cloud platforms. The integrated GPS also helps to reduce the number of modules for integration. Since NB-IOT coverage is sparse and not comprehensive, the data communication component is switched to a newer *SIM7670SA LTE/GPS transceiver* to ensure seamless coverage. The Platform Layer handles the data storage, processing, and analytics. The data is logged on the local device, i.e., a USB pen drive, in CSV format. The Application Layer translates raw data into actionable insights for interactive and real-time monitoring. The GPS data is captured as NMEA sentences, which are used to derive longitude, latitude, altitude, heading, speed, etc. The real-time results from the GPS and MPU6050 will then be transmitted via the MQTT application protocol. Figure 1 (b) shows the developed system used to experiment with this study.



(a) System architecture (b) The developed system

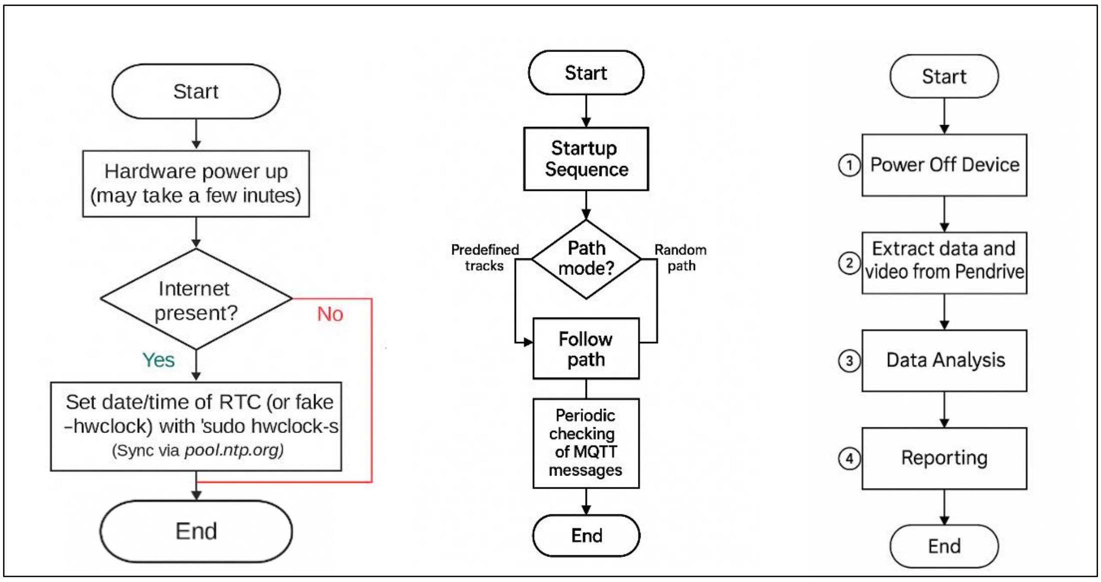
FIGURE 1: The IOT system developed

2.3 Data Collection

The developed system shown in Figure 1 (b) is used in this study. The system is started up and connected to the Internet to set the system's date and time via the Network Time Protocol (NTP). Then, the system is attached to the motorcycle's basket, and the rider is asked to follow the pre-planned route (path). The rider is asked to ride along a pre-planned route that includes various road geometries, such as curves and slopes, and different road classifications, such as highways, arterials, collector/distributor roads, and local roads. During the ride, the system status is received via MQTT messages. At the end of the journey, the system is shut down, and the data, stored in csv format, is extracted from local storage (pen drive). Figure 2 shows the data collection process, while Figure 3 shows the data format. Currently, about 100 riders have participated in this study.

2.4 Data Extraction

We define the 3D axes for the rider as follows: x forward, y left, and z up. This corresponds to the intermediate axis system specification defined in ISO 8855:2011 sec 2.13 (ISO 8855, 2011). We then define yaw (ψ) as counterclockwise rotation about the z axis, pitch (θ) as counterclockwise (CCW) rotation about the y axis, and roll (ϕ) as counterclockwise rotation about the x axis, as shown in Figure 4.



(a) Startup (b) Riding (c) Post processing

FIGURE 2: Data collection process

CSV file
datetime,Lat,Lon,Alt,Speed,Course,CGNSSINFO,vehicle,temperature,yaw,pitch,roll,Ax,Ay,Az,Gx,Gy,Gz 2025-05-11 22:02:09 +0800, N, E,0,0,0,['OK'], ybike1, 33.50352941176471, 44.994563320444726,1.1150503586650666,2.6080366048539467,- 0.01806640625,0.042236328125,0.92724609375,-0.3053435114503817,0.05343511450381679,- 0.03816793893129771

FIGURE 3: MQTT message data format

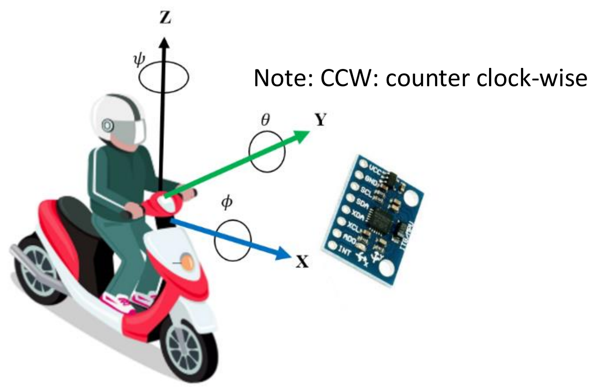


FIGURE 4: Roll, Pitch, and Yaw along the XYZ axis

The acceleration data collected from the accelerometer is used to compute and estimate the yaw (ψ), pitch (θ), and roll (ϕ). Define the accelerations as a = [a_x, a_y, a_z],

$$\phi = atan2(a_y, a_z) \quad (1)$$

$$\theta = arcsin(-a_x, \sqrt{a_y^2 + a_z^2}) \quad (2)$$

The yaw is estimated as follows:

$$\psi \approx \frac{atan2(a_y, a_z)}{v} \quad \text{from lateral specific force, OR} \quad (3)$$

$$\psi \approx \frac{g \tan\phi}{v} \quad \text{steady, level turns, without sideslip} \quad (4)$$

where *v* is the forward speed of the vehicle.
Note: Yaw cannot be computed with the accelerometer alone.*

2.5 Data Analysis

There is a lot of collected data that can be analyzed to understand various behaviors of the riders. In this study, the acceleration/braking behaviour is investigated and modelled. The longitudinal acceleration data (A_y) is extracted as mentioned in Section 2.4 and then smoothed to remove noise. The exponential moving average (EMA) is adopted to smooth the data. EMA is a widely used technique for smoothing time-series data, which gives exponentially greater weight to more recent data points. This makes EMA more responsive to recent changes or fluctuations, enabling it to capture trends more dynamically without overreacting to short-term noise. The degree of weight decrease is determined by a parameter known as the smoothing factor (α), which typically ranges between 0 and 1. A higher α value gives more importance to recent observations and produces a smoother output that closely follows the actual data.

The EMA is calculated using the following recursive equation:

$$EMA_t = \alpha \cdot a_t + (1-\alpha) \cdot EMA_{t-1} \quad (5)$$

where a_t is the longitudinal acceleration at time t;

α is the smoothing factor, $\alpha = \frac{2}{n+1}$ and $0 \le \alpha \le 1$;

n indicates the size of the data.

Clustering analysis is adopted to classify the behaviour of riders based on smoothed longitudinal acceleration data. K-means clustering is used to classify rider behaviour by partitioning the dataset to minimize variance within each cluster and maximize variance between clusters. The algorithm initializes K cluster centroids randomly and assigns each data point to the nearest centroid based on the Euclidean distance. Then, the centroids are updated by calculating the mean of the points in each cluster. These steps are repeated until the convergence factor is met. In this study, we define K=3, with riders classified into three categories: normal, aggressive braking, and aggressive acceleration. The algorithm stops when the movement of centroids between iterations falls below a set tolerance of <0.10.

3. RESULTS AND DISCUSSION

As the study is still ongoing, a preliminary analysis of the longitudinal acceleration data of one of the riders is presented to show the applicability of the IOT system and analysis method. Figure 5 shows the location where the rider performs the three types of actions, i.e., normal riding, aggressive braking, and aggressive acceleration.

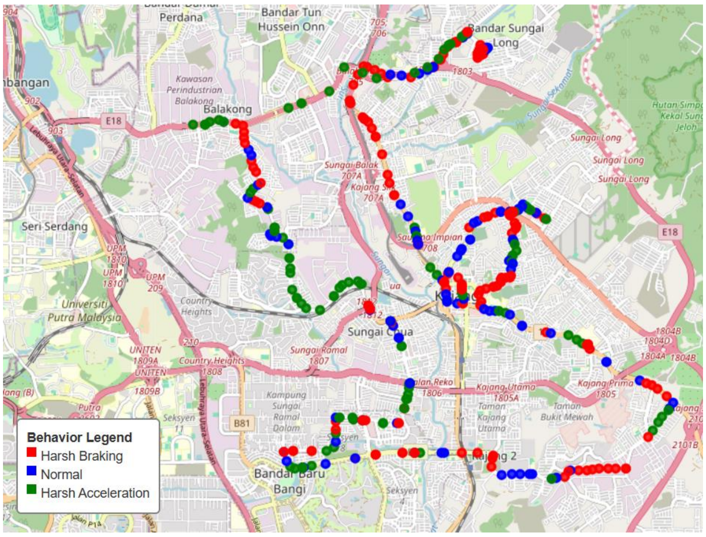


FIGURE 5: Results - driver behaviour classification

It is observed that multiple occurrences of harsh braking (red markers) occur along a specific stretch of the road. These events are concentrated on several consecutive points, suggesting repeated deceleration over a short distance. This pattern indicates that the rider is encountering conditions that require a sudden speed reduction rather than maintaining smooth, continuous movement. The clustering of harsh braking points corresponds to areas likely associated with curves and intersections. In such geometries, riders typically reduce speed abruptly due to visibility constraints, cross-traffic, or tighter turning radii. Besides, repeated speed adjustments occur during cornering on curved alignments. Harsh accelerations are observed at locations after the intersections. While on the straight sections, normal riding behaviour is observed.

4. CONCLUSION

This study demonstrates the feasibility of using an IoT-based system to support NRS for motorcyclist behaviour analysis. By integrating affordable components such as the MPU6050 IMU, Neo-M6 GPS, NB-IoT SIM7000E, and Raspberry Pi 3, the system effectively captures key behavioural metrics, including acceleration, angular velocity, orientation (pitch, roll, yaw), and location data in real time. The layered architecture enables reliable data acquisition, transmission, and processing under real-world riding conditions. Pilot testing confirms the system's applicability. Compared to traditional methods, the IoT-based approach offers a more detailed and context-sensitive means of studying rider behaviour and risk exposure. It presents a practical, scalable solution to improve road safety research and inform targeted interventions for motorcyclist crash prevention.

ACKNOWLEDGEMENT

The authors would like to thank all members of the ASEAN NCAP for their valuable input and strong cooperation throughout this work. This research was supported by the ASEAN NCAP ANCHOR V Grant.

REFERENCES

Abdul Manan, M.M. & Várhelyi, A. (2012). Motorcyclists' road safety problems in Malaysia. IATSS Research. 36(1), pp.30–39.

Abdul Manan, M. M., Mohamad Khaidir, N., & Mohamed Jamil, H. (2020). Factors associated with red-light running among motorcyclists at signalised junctions in Malaysia. Transportation Research. Part F, Traffic Psychology and Behaviour. 73, 470–487.

Cho, E., et al. (2023). Derivation of riding risk precursors using 100 delivery scooters' NRS data (Korea). Accident Analysis & Prevention. 194, 107384.

Espié, S. (2013). Data collection and processing tools for naturalistic study of powered two-wheelers users' behaviours. Accident Analysis & Prevention. 60, pp.393–403.

Halim, W., Sartika, E.M., Jarden, J.J. & Iridiastadi, H. (2025). Design and implementation of an instrumented motorcycle for a naturalistic driving study in Indonesia. Engineering Proceedings. 84(1), p.16.

ISO 8855 (2011). ISO 8855:2011 Road vehicles — Vehicle dynamics and road-holding ability — Vocabulary. International Standards Organization. Retrieved from https://www.iso.org/

Jarod, L. (2024), Putting the brakes on biker road deaths. The Star. Retrieved from https://www.thestar.com.my/metro/metro-news/2024/11/26/putting-the-brakes-on-biker-road-deaths

Kim, D., Kim, H., Lee, S., Lee, Q., Lee, M., Lee, J., & Jun, C. (2024). Design and Implementation of a Two-Wheeled Vehicle Safe Driving Evaluation System. Sensors (Basel, Switzerland), 24(14), 4739.

MIROS (2022). Road Safety Annual Report 2022. MIROS Publications.

MOT (2022). Malaysia Road Safety Plan 2022-2030. Ministry of Transport Malaysia. Retrieved from https://www.mot.gov.my

Raspberry Pi Ltd. (n.d.). Operating system images. Raspberry Pi Holdings. Retrieved from https://www.raspberrypi.com/software/operating-systems/

Williams, S. (2014). The MSF-100 motorcyclists naturalistic study: Preliminary results. In: New York Highway Safety Annual Fall Symposium, Lake Placid, NY, October 2014. Motorcycle Safety Foundation.