


Article

Improving Autonomous Vehicle Controls and Quality Using Natural Language Processing-Based Input Recognition Model

Mohd Anjum ¹ and Sana Shahab ^{2,*} 

¹ Department of Computer Engineering, Aligarh Muslim University, Aligarh 202002, India

² Department of Business Administration, College of Business Administration, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

* Correspondence: sshahab@pnu.edu.sa

Abstract: In contemporary development, autonomous vehicles (AVs) have emerged as a potential solution for sustainable and smart transportation to fulfill the increasing mobility demands whilst alleviating the negative impacts on society, the economy, and the environment. AVs completely depend on a machine to perform driving tasks. Therefore, their quality and safety are critical concerns for driving users. AVs use advanced driver assistance systems (ADASs) that heavily rely on sensors and camera data. These data are processed to execute vehicle control functions for autonomous driving. Furthermore, AVs have a voice communication system (VCS) to interact with driving users to accomplish different hand-free functions. Some functions such as navigation, climate control, media and entertainment, communication, vehicle settings, vehicle status, and emergency assistance have been successfully incorporated into AVs using VCSs. Several researchers have also implemented vehicle control functions using voice commands through VCSs. If a situation occurs when AV has lost control due to malfunctioning or fault in the installed computer, sensors and other associated modules, driving users can control the AV using voice notes to perform some driving tasks such as changing speeds, lanes, breaking, and directing the car to reach a safe condition. Furthermore, driving users need manual control over AV to perform these tasks in some situations, like lane changing or taking an exit due to divergence. These tasks can also be performed with the help of voice commands using VCSs. Therefore, finding the exact voice note used to instruct different actuators in risk situations is crucial. As a result, VCSs can greatly improve safety in critical situations where manual intervention is necessary. AVs' functions and quality can be significantly increased by integrating a VCS with an ADAS and developing an interactive ADAS. Now, the driver functions are controlled by voice features. Therefore, natural language processing is utilized to extract the features to determine the user's requirements. The extracted features control the vehicle functions and support driving activities. The existing techniques consume high computation while predicting the user command and causing a reduction in the AVs' functions. This research issue is overcome by applying the variation continuous input recognition model. The proposed approach utilizes the linear training process that resolves the listening and time-constrained problems and uncertain response issues. The proposed model categorizes the inputs into non-trainable and trainable data, according to the data readiness and listening span. Then, the non-distinguishable data were validated by dividing it into the linear inputs used to improve the response in the AVs. Thus, effectively utilizing training parameters and the data decomposition process minimizes the uncertainty and increases the response rate. The proposed model has significantly improved the exact prediction of users' voice notes and computation efficiency. This improvement enhances the VCS quality and reliability used to perform hand-free and vehicle control functions. The reliability of these functions ultimately improves the safety of AVs' driving users and other road users.

Keywords: autonomous vehicle; natural language processing; variation continuous input recognition; uncertainty analysis



Citation: Anjum, M.; Shahab, S. Improving Autonomous Vehicle Controls and Quality Using Natural Language Processing-Based Input Recognition Model. *Sustainability* **2023**, *15*, 5749. <https://doi.org/10.3390/su15075749>

Academic Editors: Cornelia Mairean, Danijela Baric and Grigore M. Havarneanu

Received: 14 December 2022

Revised: 21 March 2023

Accepted: 23 March 2023

Published: 25 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the current era of development, there has been a steady paradigm shift in transportation and mobility. This paradigm shift is changing everything, from the fuel used to how vehicles are driven [1]. In this paradigm shift, many novel technologies have emerged to develop intelligent transportation and sustainable urban mobility. These technologies are progressively focusing on making vehicles autonomous and capable of communicating and cooperating [1]. As a result, in a prolonged period, substantial radical transformations are expected in mobility as a service in the prospect of future mobility solutions for smart and sustainable development. In addition, urban regions will have a faster and higher level of development, and cities will be expected to convert into smart cities where the vehicles will communicate with the urban infrastructure and driving users will be able to interact with them [2].

Autonomous vehicles (AVs) have begun to appear on city roads. These vehicles have a significant role to play in the future of sustainable and smart transportation for urban regions [3]. Sustainable and smart transportation systems significantly mitigate the adverse effects of urban development on the environment, economy, and society [4]. The widespread adoption of AVs around the globe can decrease environmental degradation by controlling emissions and minimizing energy consumption. It can also provide economic and social benefits by improving the efficiency, safety, and accessibility of transport services [5]. They have unique capabilities and are equipped to provide a safe travel mode by eliminating human driving errors [6].

Contrary to humans, AVs perform driving tasks tirelessly without any distractions. Autonomous driving has recently moved from the “may be possible” domain to “has happened practically.” Beyond the safety, security, and entertainment for the driving users, AVs also contribute one step forward in smart and sustainable development. They are an emerging technology that provide better services and performance for users via automatic driving skills. AVs do not require humans to drive them. The backbone of AVs’ development is the revolutionary growth in sensors and communication technologies [7]. Various types of sensors and communication modules, namely radio detection and ranging, light detection and ranging, ultrasonic, camera, and global navigation satellite systems, are used in AVs to perceive the surrounding environment and gather related information [8]. Powerful computers with specialized software, machine learning (ML) systems, artificial intelligence models, complex algorithms, and hard-coded rules are used to process the captured data and make logical decisions to accurately perform the driving task in a complex environment like humans [9]. After processing, the computer directs the actuators to act for uninterrupted driving. This system of self-driving is called ADAS.

Besides ADAS, which comprises the sensors and communication modules mentioned above, a VCS is used in AVs to interact with them [10]. Driving users of AVs communicate with VCSs to perform a variety of hand-free functions. Some functions include navigation (setting a destination, changing routes, and searching for points of interest); climate control (temperature change, fan speed, and airflow); media and entertainment (operating infotainment systems such as changing the volume, skipping tracks, or switching between radio stations); communication (phone calls or sending and receiving emails); vehicle settings (headlight and windshield wiper controls); vehicle status (vehicle’s current speed, fuel level, door and window locks and other status information); and emergency assistance (calling for help or requesting roadside assistance have been successfully incorporated into AVs using VCSs) [11]. Several researchers have also implemented vehicle control functions using voice commands through a VCS. Several researchers have also implemented vehicle control functions, including turn signals, gear selection, engine control lane changes, or taking an exit due to divergence using voice commands through VCSs [11]. The VCS plays a critical role in the safety of driving users in situations where the AV has lost control due to malfunctioning or faults in the hardware of the installed computer, sensors, and other associated modules. In this uncertain and risky situation, driving users can control the AV using voice notes to perform driving tasks such as changing speeds, directions,

and lanes, as well as braking, to reach a safe condition. Furthermore, driving users need manual control over AVs to perform these tasks in some situations, like changing lanes or taking an exit due to divergence. These tasks can also be performed with the help of voice commands using a VCS. Therefore, it is crucial to determine the exact voice note used to instruct different actuators in a risky situation. The current VCSs face various challenges in interpreting actual commands from voice notes due to listening issues and time-constrained and uncertain response problems. The study addresses these issues using an input recognition model based on natural language processing (NLP). Overall, the proposed model improves AV controls, directly enhancing the safety of driving users in risky situations.

AVs provide various functions and processes for users, ensuring their safety and security. The identification, classification, detection, and analysis processes are widely used in AVs to provide appropriate services for users [12]. Identifying voice commands is a complicated task in AVs, which is necessary for various methods and analysis processes. A VCS is used in AVs to find the exact command from the voice note, which is used in the analysis process [13]. VCSs provide an accurate set of voice notes, ensuring a high accuracy rate in the data processing system. VCSs use intelligent techniques to learn how to interact with people and recognize users' voice commands while travelling in AVs, providing uninterrupted and accurate user services [14]. A VCS is installed in AVs to capture users' voice commands and securely store data for later use. The voice user interface method is used in AVs, utilizing a user mental model to identify the exact voice commands [15]. Artificial intelligence and the big data analysis process are used in AVs to fetch related data for identification. The user mental model is used to determine how the user thinks and find the exact meaning of voice commands to perform tasks or services in AVs [7].

NLP is an interactive process between computers and human language. It is a branch of computer science that provides an accurate understanding of data for analysis. Human languages are separated into fragments to find the grammatical structure and give the correct meaning of the sentence, which plays a vital role in the data processing system [16]. When analyzing a large amount of data, NLP provides a better set of data or a way for computers to reduce latency. AVs are an emerging technology that offer better services and performance for users via automatic driving skills [17]. AVs ensure user safety by providing various services and functions to enhance the system's feasibility. AVs use NLP to offer an accurate communication process for users, reducing the rate of AV accidents [18]. NLP provides features such as text format, text structure, and sentence size to improve classification and identification rate accuracy. NLP determines the format of text and structure to identify the text's actual meaning and content, providing an accurate set of data for the data processing system in AVs. NLP uses a process called "knowledge discovery" that identifies the meaning of the text and improves the feasibility of AVs [19,20].

ML is a subset of artificial intelligence used to improve the accuracy rate of the prediction, detection, and analysis processes in various fields. ML techniques are used in AVs to enhance the performance of the system [21]. NLP provides a better interaction process between computers and humans, increasing the system's feasibility. ML-based NLP methods are widely used in AVs to improve the accuracy rate of identifying the exact meaning of a user's communication [22]. A neural network is used in NLP to identify the pattern and structure of the text, which produces an accurate set of data for the analysis process. The identified patterns are converted into vectors using a network, which produces the actual meaning of text and sentences for AVs [23]. The produced set of data is used in various AV processes, which improves users' safety and reduces the accident rate. Combined ML techniques are also used in the NLP process, providing an exact dataset for the analysis process. Support-vector machines and deep-learning algorithms are combined to form a new technique to perform NLP in AVs, enhancing the system's performance [24,25].

The sustainability of a transportation system depends on its users' safety, with human life being the most valuable resource. Ensuring the safety of individuals is a priority and

should be a fundamental aspect of any sustainable transportation system. The ultimate objective of a sustainable and safe road transportation system is to eradicate fatalities, severe injuries, and permanent harm by systematically addressing the issue and reducing the inherent hazards of the entire road transportation system [26]. Road user behaviour is a crucial aspect of transportation and mobility as it is considered the main contributing factor in most road crashes. Therefore, understanding and addressing the behaviour of road users is crucial in reducing the number of accidents and fatalities on the roads [27]. Road user behaviour involves understanding the human factors, such as physical [28], psychological [29], cognitive [30], infrastructure [31], climate [31], and technological factors influencing road users' actions [32]. Many studies have shown that driver personality traits, such as age, gender, sleeping hours, working hours, reckless driving, distracted driving, and road user education, are associated with increased risk-taking behaviour on the road [33].

Additionally, cognitive factors, such as driving stress and decision-making abilities, have been found to play a role in road user safety [34]. Infrastructure and climate conditions include the design and infrastructure of the road network [35] and weather conditions [31]. Technological factors refer to advanced technologies promoting safe behaviour, such as ADASs [36,37], navigation systems [38], and AVs [39]. Nowadays, researchers have focused on the relationship between technology and road user behaviour [38]. With the increasing prevalence of ADASs and AVs, there is a growing need to understand how these technologies can promote safe road user behaviour. Many studies have investigated the effectiveness of ADASs and AVs in reducing the crash rate and accidents and improving driver performance [37,40]. Automated driving has the potential to revolutionize road transportation by increasing safety, improving traffic flow, and providing mobility for all [40]. Furthermore, AVs have eliminated the impact of humans and related factors by removing human involvement in driving tasks [41]. Therefore, the study has proposed a natural language processing-based input recognition model to improve AV controls and quality, which ultimately contributes to road user safety.

In the context of AVs, NLP can be used to improve the accuracy of voice commands given to the vehicle. Therefore, applying NLP reduces errors and misunderstandings and improves the AVs' overall functionality. ML algorithms train the vehicle to recognize specific voice commands and make decisions based on those commands. Consequently, ML algorithms improve the safety of AVs by reducing the risk of accidents caused by human error. Research in the field of VCS application has focused on combining NLP and ML techniques to improve the functionality and safety of AVs. Many studies have proposed methods for enhancing the VCS in AVs, utilizing NLP techniques for input recognition and ML techniques to enhance the performance of AVs. For example, some studies proposed a real-time traffic reporting system using NLP for social media networks, and others proposed a visualizing natural language interaction for a conversational in-vehicle information system. Overall, using NLP and ML techniques in AVs can improve the accuracy of voice commands, enhance AVs' performance, and increase AVs' safety by reducing the risk of accidents caused by human error.

VCSs in AVs play a crucial role in improving road user safety. These systems allow for hand-free and vehicle control functions, reducing potential risks. Using NLP and ML techniques, the system can interpret and respond to spoken commands, such as navigation instructions, climate control adjustments, or infotainment system controls, vehicle control functions and driving functions. Additionally, in the case of a malfunction or failure in the vehicle hardware, the VCS can act as a fail-safe mechanism, allowing the driving users to take control of the vehicle using voice commands to safely bring the vehicle to a stop or navigate to a safe location. As a result, VCSs can greatly improve safety in critical situations where manual intervention is necessary. The integration of VCSs with ADASs in AVs significantly contributes to road user safety by providing a fail-safe mechanism. The proposed variation continuous input recognition model (VCIRM) is a novel approach for continuously interpreting spoken commands or input. It allows for variations in how

a command is spoken, such as accents, speed, and phrasing. In contrast to traditional input recognition models, which may only recognize specific, pre-determined phrases or commands, the proposed VCIRM can more accurately understand and respond to spoken commands, even if they are spoken differently than originally anticipated. Additionally, it increases the flexibility of the system, allowing it to respond to a wider range of user inputs. This model is frequently used in NLP and speech recognition systems, such as those used in AVs, to improve the accuracy of voice commands and enhance the performance and safety of the vehicle.

The paper is divided into five sections. Section 1 provides the research background and rationale behind the research. Section 2 reviews the existing literature in the field and highlights the gap in the current state of the art that the proposed model aims to address. Section 3 describes the design and implementation of the proposed VCIRM, as well as the techniques used to train and evaluate it. Section 4 presents the results of experiments conducted to assess the performance of the proposed model, comparing it to existing models and discussing the results. Finally, Section 5 summarizes the key findings of the study and highlights the importance of the proposed model in the context of driving users' safety.

2. Related Works

In recent years, the field of AVs has seen significant research focused on improving these vehicles' functionality and safety through NLP and ML techniques. NLP is used to extract meaning and structure from human language, while ML involves using algorithms and statistical models to analyze large amounts of data and make predictions. These techniques can be combined to improve the accuracy of voice commands given to AVs and enhance the performance and safety of these vehicles. This literature survey aims to provide an overview of the various studies conducted in this field, highlighting the recent developments and current state of research on NLP and ML for AVs.

Wan et al. [42] introduced an automated NLP-based framework (ANLPF), which is a real-time traffic reporting system using NLP for social media networks. The proposed method performs a text-mining process to find the exact meaning and content of the text, providing accurate data for drivers and users. A question-answering model is used to extract information or data for users, which plays a vital role in identifying traffic flow on roads. The proposed traffic reporting system is more accurate than other methods in regard to giving users information. Braun et al. [43] proposed visualizing natural language interactions for a conversational in-vehicle information system. The proposed method improves the speech-based interaction process in the in-vehicle system. A certain set of keywords is given to understand the exact content of the text, which enhances the interaction process for the users. The attractiveness of the interface is increased by using icons and symbols that provide accurate detail about the interaction process. The proposed method improves the visualization of the interaction process, which increases the accuracy rate in the prediction and detection processes. Solorio et al. [44] introduced an off-the-shelf home automation component-based semi-autonomous utility vehicle. The proposed method is a voice-activated automated system that uses hardware and software elements for interaction. The proposed approach is mostly used in web and smart applications to enhance control and command over vehicles, improving the system's performance. A speaker and voice recognizer are used in a vehicle to provide accurate information and services for users.

Choi et al. [45] developed an active-beacon-based driver sound separation (ABDSS) system using the concept of an active beacon for AVs. Voice command plays a vital role in this system, which provides actual and optimal voice commands for interaction and service processes. The proposed system would separate the driver's voice from other voice commands so that services in AVs would be more accurate. Voice signals are identified using a distinguishing process that enhances the efficiency and feasibility of the system. Riaz et al. [46] introduced an emotion-inspired cognitive agent scheme for

spectrum mobility in cognitive-radio sites. The proposed scheme improves the efficiency of spectrum mobility using the fear factor. The proposed scheme increases mobility's speed and accuracy rate using the fuzzy logic algorithm. Experimental results show that the proposed agent increases the system's performance and spectral mobility rate. Saradi et al. [47] proposed a voice-based motion control scheme for the robotic vehicle using a visible light-fidelity communication process. In this system, an artificial neural network is trained for the interaction data to control the motion of the proposed vehicle. Here, the light fidelity process increases data bandwidth and efficiency, providing better user service and communication. The proposed scheme significantly improves the accuracy rate in the interaction process, enhancing the system's feasibility and reliability.

Sachdev et al. [48] introduced a voice-controlled AV using the Internet of Things to determine the user's exact location, position, and direction via a voice-controlled remote sensing system. The Internet of Things provides necessary information about an AV using surveillance cameras and a global positioning system. The AV follows the user's voice commands, reducing accidents and latency rates in providing services. The proposed method improves the overall performance and efficiency of the system. Ni et al. [49] proposed a domain-specific natural language information brokerage for AVs. The proposed method works as a task helper, providing necessary services for users at the appropriate time. A question-answering mechanism is used in the proposed approach to utilize essential data to provide accurate user service. The proposed method improves accuracy in delivering relevant and precise services that are of high quality. Zhang et al. [50] introduced a lightweight vehicular voice cloud evaluation system (LVVCES) for AVs. First, voice signals are sent to the cloud to find the user commands needed for providing services for the users. The tester is used to identify the optimal solution and data for the analysis process, which reduce unwanted problems and threats in the communication process. The proposed system increases the overall quality of experience of the AV, enhancing the system's performance. Katsikeas et al. [51] proposed a vehicle modelling and simulation language for AVs. The proposed method is used to provide better security for vehicles from vehicular cyberattacks, and it uses a vehicle-to-vehicle (V2V) approach to improve AVs' communication and authorization processes. The proposed method is also used for risk management and threat modelling for AVs, which increase the system's efficiency.

Wang et al. [52] introduced a distributed dynamic route guidance system for a cooperative vehicle infrastructure using short-term forecast data. Short-term forecast data are used in the distributed dynamic route guidance system for the prediction and detection processes. The proposed method reduces threats and problems in the prediction and analysis processes, which increase the system's performance. The results of the experiments show that the proposed guidance system makes a cooperative vehicle infrastructure system more efficient and possible. Asmussen et al. [53] proposed a socio-technical AV model using ranked-choice stated preference data. The proposed model is used to determine the AV's mobility rate, speed, accuracy, and control rate for the users. The proposed socio-technical model provides an optimal dataset for further processing and operation in an AV. The socio-technical model determines users' precise voice and text commands to provide services.

Zheng et al. [54] introduced a new V2V communication process for AVs. The proposed method promotes cooperative lane changes in a V2V communication system, which enhance the communication process for the users. In lane changes, the collision trigger time is used to improve the communication process in AVs. Experimental results show that the proposed method improves the performance and safety of users from attackers. Totakura et al. [55] focused on developing self-driving cars using convolutional neural networks and identifying and addressing potential drawbacks. The developed model for self-driving cars was trained using data from the Asphalt-8 game, while a separate convolutional neural network model for voice-command prediction was trained with the voices of a child, man, and woman. The accuracy of both models was found to be 99%, and they were tested on the same game for optimal results. This research demonstrates

the effectiveness of using a convolutional neural network model in self-driving cars and highlights the importance of addressing drawbacks to ensure safe and sustainable road user behaviour.

3. Proposed Variation Continuous Input Recognition Model

Variations in the AV interactive gesture and voice control processing were experienced with the safety and driver assistance. The VCS system uses automatic systems trained and loaded with several pre-defined comments and functions. These functions instruct the driver to perform the safety driving actions, which ensure overall safety. Amid the challenges in interactive voice systems, NLP features in AVs use quality control and data availability to identify user requirements and satisfy different drivers who use driving support. The driving supports of users from adaptive cruise control, autonomous emergency braking, electronic stability control, blind-spot detection, V2V communication, vehicle guidance system, voice recognition, and control require distinguishable services. Therefore, regardless of the interactive system voice input and detection of the vehicles, data availability of indistinguishable and non-distinguishable data for training is a prominent deciding factor. Figure 1 illustrates the schematic diagram of the proposed VCIRM.

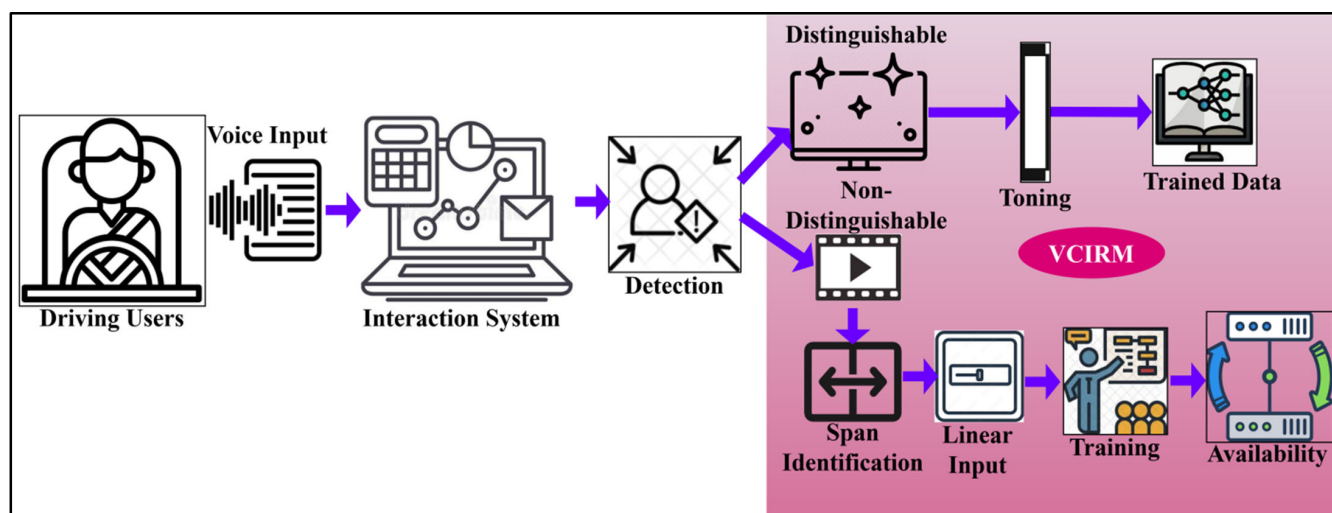


Figure 1. Schematic diagram of variation continuous input recognition model.

The proposed VCIRM focuses on the listening span and data readiness of available data toning through a linear training process. In this approach, internal controls or external driving supports are administrable for driving users and their trainable and non-trainable data based on the response lapses. AV driving users can access interactive voice input by detecting perfect voice recognition, identifying the user's requirements, and responding using NLP. The proposed VCIRM model operates between the vehicles and driving users. In this model, distinguishable and non-distinguishable data for the available internal controls and driving support are feasible for achieving response lapses for the different users and vehicles. This voice input recognition model also aims to provide split-less responses and maximize data availability. The proposed model operates in two forms, distinguishable and non-distinguishable, concurrently. The non-distinguishable data differ from trainable and non-trainable data to handle different internal controls or external driving supports, as shown in Figure 1. The introducing operations of the interactive voice input of AVs driving users are keen about the objective function shown in Equations (1a,b).

$$\underset{n \in u}{\text{maximise}} \quad D_n \quad \forall Rq = Rs \quad (1a)$$

Such that,

$$\underset{m \in nd}{\text{minimise}} \ rT_m = T_{Rq} - T_{Rs}, \text{ where } \underset{n \in u}{\text{minimise}} \ d_n \ \forall \ n \in Rq \quad (1b)$$

In Equations (1a,b), the variables D_n , Rq , Rs , d_n represent interactive voice input detection of a n^{th} driving users u , requirements, responses, and distinguishable data, respectively. In the next consecutive representation, the variables rT_m , T_{Rq} , and T_{Rs} denote response time, user requirement accepting time, and input responding time. The third objective of this technique is to minimize the distinguishable data using the variable $d_n \ \forall \ n \in Rq$. If $u = \{1, 2, 3, \dots, u\}$ denotes the set of driving users, then the number of voice input detection in the user requirement accepting time is $Rq \times T$, whereas the user requirements are $u \times Rq$. Based on the overall AV driving users of $u \times Rq$, $T \times Rq$ are the admissible process for detection.

Voice input detection and perfect recognition processes are reliable using toning and training *et* of the upcoming data. In this research, toning and training data variations are essential to identify non-trainable additional data. The demanding requirement is the linear input (L_n) of the n driving users; the remaining time needed for distinguishable data is the helping factor for improving the training rate. The detection of the voice input data assigned for the available n is functional using a linear learning process. Later, depending upon the detection of the interactive voice system, the non-distinguishable process is the augmenting feature. From this detection process, listening span and data readiness are the prevailing instances for determining various constraints. The pre-modelling of data and the availability requirements for training are essential in the following section.

3.1. Case 1: Distinguishable Data Detection

In this distinguishable data, the detection of $(Rq \times T)$ for all n driving users based on L_n is the considering factor. The distinguishable data detection process is illustrated in Figure 2.

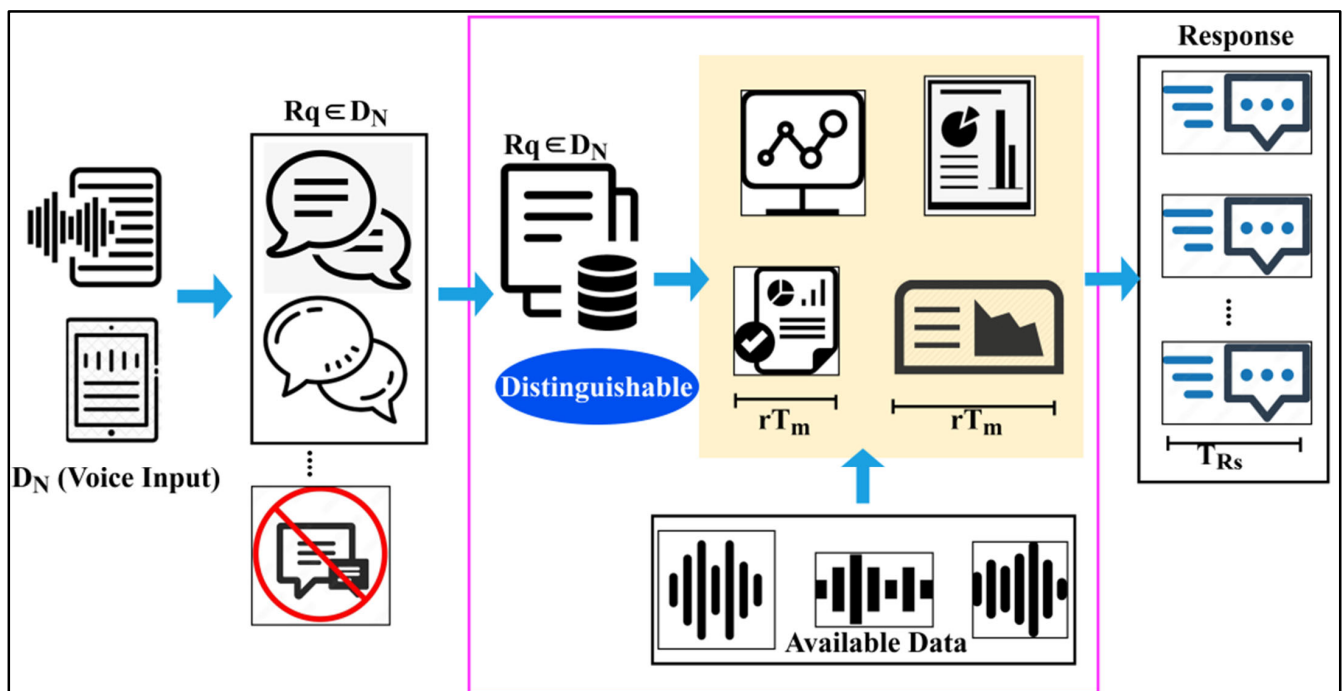


Figure 2. Distinguishable data detection process.

Via indistinguishable data processing, the common interactive inputs are segregated from the unfamiliar (unrecognisable) inputs. The $R_q \in D_N$ and $R_q \in D_N$ and ρ_t are

differentiated. This differentiation is performed to accept $R_q \forall rT_m$ for the toning process. From this processing, T_{RS} is the time required for responding to inputs. The ratio of $R_q \in D_N$ from R_q is required for the consecutive classification of D_N , as shown in Figure 2. The probability of distinguishable data (ρ_d) consecutively is given in Equations (2a,b).

$$\rho_d = (1 - \rho_t)^{n-1} \forall n \in T \quad (2a)$$

where,

$$\rho_t = \left(1 - \frac{R_q \in n}{R_q \in T}\right) \quad (2b)$$

From Equations (2a,b), the sequential detection of voice input follows the constant probability of n such that there are no uncertain responses. Therefore, et is as estimated in Equations (1a,b). Hence, the detection of distinguishable user requirements for ρ_d follows Equation (3).

$$Detection(n) = \frac{1}{|Rs - Rq + 1|} \cdot (\rho_d)_n \forall n \in T \quad (3)$$

However, the distinguishable data detection for n , as in Equation (3), is valid for both the conditions $(u \times Rq)$ and $(T \times Rq)$ handling to ensure time-constrained listening responses. With the converging process of perfect recognition T to reduce the problem of the constraint $(u \times Rq) > (T \times Rq)$, the distinguishable data is descriptive using detection or perfect recognition. Therefore, the identifiable constraint is $u > T$, and ρ_t denotes the trainable data, which is less to satisfy Equation (1a,b). The contrary output in this Case 1 is the prolonging ρ_t . Therefore, the response time results in a lower response rate.

3.2. Case 2: Non-Distinguishable Data Detection

In a non-distinguishable data detection process, the uncertain condition of $u > T$ is high. Hence, the internal control/external driving support of users is time-constrained. In addition to the constrained time of n , the trainable and remaining information are considered metrics for this case. The non-distinguishable data detection process is presented in Figure 3.

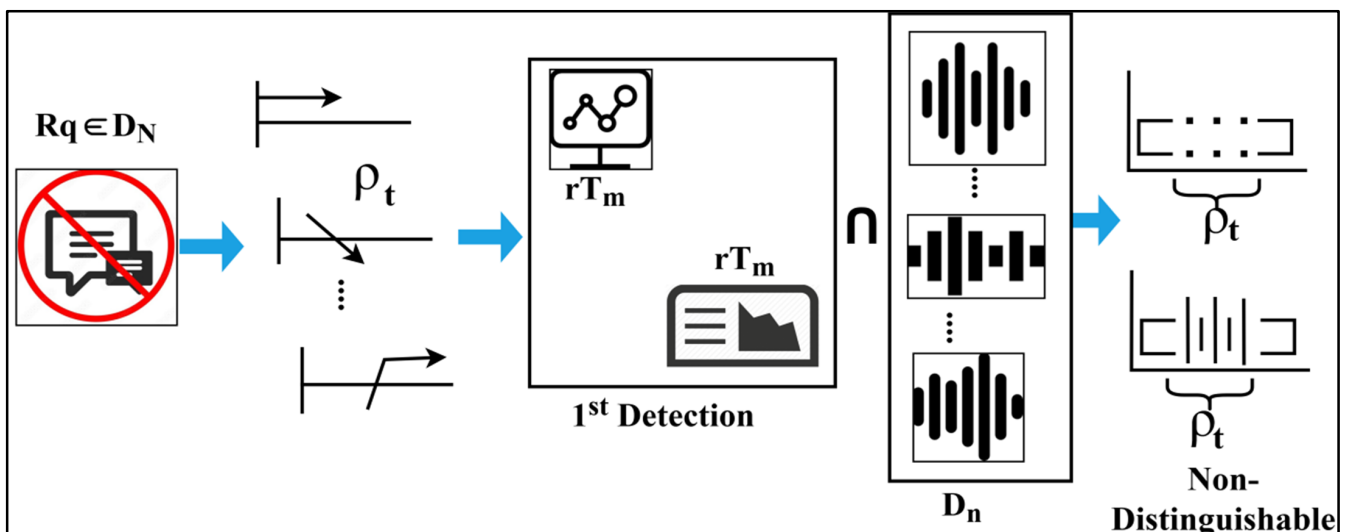


Figure 3. Non-distinguishable data detection process.

The $R_q \in \rho_t$ is identified as non-distinguishable, from which the non-consecutive sequences are segregated. Based on the ρ_t , rT_m and D_N are cross-validated for extracting $R_q \in \rho_t$. This extraction is performed to prevent an anonymous D_n , a distinct interval (before classification). Therefore, for the varying L_{span} , the process is unanimously pursued,

preventing uncertainty, as presented in Figure 3. The probability of non-distinguishable data (ρ_{Nd}) is given by Equations (4) and (5a,b).

$$\rho_{Nd} = \frac{\rho_d \cdot \text{Detection}(n) \times \left[(Rs - Rq)\rho_t - \left(\frac{Rs - Rq}{n} \right) \frac{rT}{T_{Rq}} \right]}{O_p(a) \cdot n} \quad (4)$$

where,

$$O_p(a) = \int_0^t rT^{t-1}(1 - rT)^{t-1} d(rT) \quad (5a)$$

Such that,

$$O_p(a) \in \text{Detection}(n) = \int_1^{Rq} rT^{t-1} \cdot \frac{\rho_t}{T_{Rq}} (1 - \rho_d)^{t-1} d(Rq) \quad (5b)$$

Based on the above Equations, the variable $O_p(a)$ denotes the interactive voice input detection operation for t . For all the detection processes, the uncertainty in assigning information to n is the training data problem. As in the above constraint, voice detection requires a greater response time, thereby increasing the training rate.

According to the analysis of Cases 1 and 2, the variation condition of uncertainties based on Case 1, $u > T$ and N training data, and the responding time are the identifiable conditions. These conditions are addressable using linear learning to mitigate the issues through the toning process. The following section represents the toning process for the distinguishable data.

3.3. Distinguishable Data Using the Toning Process

The decision for toning (matching) distinguishable data relies on a linear learning paradigm. It supports data availability for both discrete and continuous sequences. Case 1 (continuous/distinguishable) and Case 2 (distinct/non-distinguishable) processes are toning with the resolving instances using linear input. The matching process depends on various factors for analyzing the trainable data and uncertainty probabilities during interactive systems detection. Therefore, the above cases for voice input detection are different; they follow distinguishable procedures through the toning process. The toning process for continuous and distinct identifications is represented in Figure 4.

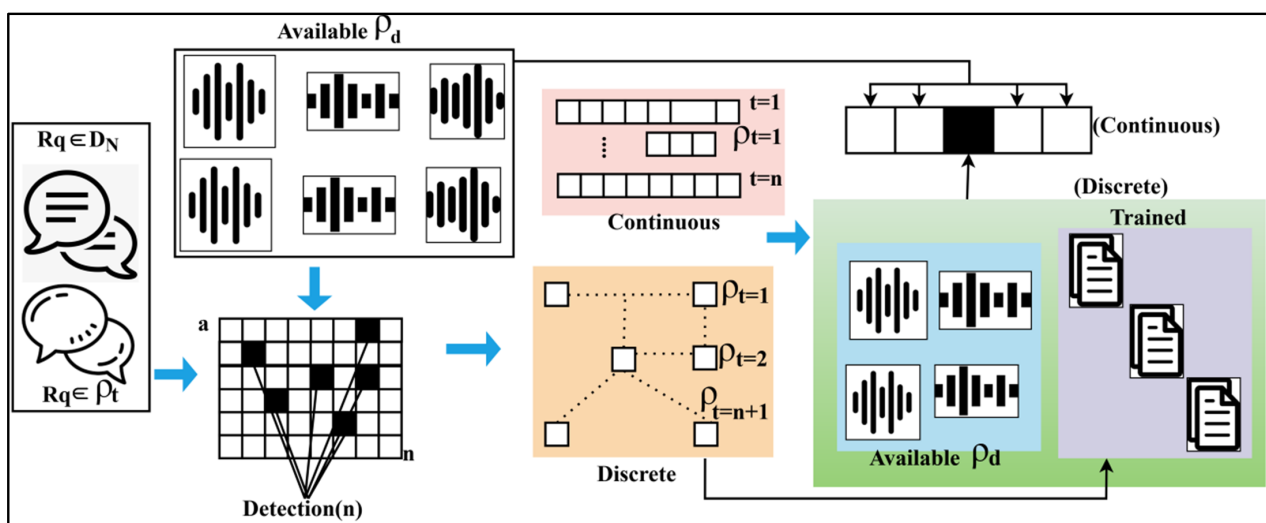


Figure 4. Toning process for continuous and distinct identifications.

In the toning process, $R_q \in D_N$ is induced for ρ_d and ρ_{v1}, ρ_{v2} data for analysis. The non- ρ_d data are trained in the $(n+1)^{th}$ instance for improving $Detection(n)$ with ρ_t . This operation is performed $\forall \rho_t = 1$ to $n+1$, such that the data availability-based validation is performed, as depicted in Figure 4. The toning is prescribed for both Cases 1 and 2 by computing the n available probability and detection of voice data for a constrained time. The first toning relies on maximum training data (T_{on}) and $O_p(a)$, as given in Equations (6a,b).

$$O_p(a, T_{on}) = \left[Rs - \left(\frac{rT}{T_{Rq}} \right) \times \frac{1}{n} \right] - Detection(n) + 1 \quad (6a)$$

Such that,

$$n = \sum_{T \in r} Detection(n)_T - (\rho_{Nd})_T \quad (6b)$$

In the computation of Equations (6a,b), the main goal is to address the linear training u and T to reduce the responding time. Therefore, the actual Rs is given in Equation (6c).

$$Rs = \max \left[\frac{\rho_d \times Rq}{Detection(n) - \rho_t \times Rq} \right] \quad (6c)$$

Therefore, the uncertainty is $\left[1 - \frac{\rho_d \times Rq}{Detection(n) - \rho_t \times Rq} \right]$, and this internal control is the responding time training instances of Rq . The excluding Rq is $[r_q * O_p(a, T_{on})]$, which is the d_n obtaining sequences. Hence, the response time is demandingly high. The remaining $Rs \forall T \in Rq$ is estimated using Equations (6a–c). Therefore, the next et is essential for detecting the remaining user requirements. In this case of distinguishable processing, n (or) $\left(n - \frac{Rq}{rT} \right)$ is the data availability irrespective of the users and vehicles. In the next section of interactive voice input detection, minimizing $d_n = \{1, 2, 3, \dots, Rs\}$ [as from Equation (6a)] is discussed to reduce training data and response lapses.

3.4. Non-Distinguishable Data Using Linear Input

The non-distinguishable data process follows either of the Rs , as in the above section. It is different for both Rs in the first instance to obtain no more n , whereas the next instance, which obtains non-trainable data as $(n - Rq)$, retains user requirements. Based on the discussion in the previous section, the detection of distinguishable data for $d_n \in Rs = \frac{(n+1)Rq}{n}$ is reliable, and it does not require lapse/response time. The listening span (L_{span}) of a T in this detection is the deciding factor, and it differs for each n , depending on the availability of processing (n_a). This time is evaluated using Equation (7) for both Rs in Equation (6a,b,c).

$$L_{span} = \begin{cases} \frac{n_a}{Detection(n)}, & \text{if } \forall Rs = Rq \\ \frac{n_a}{Detection(n)} + \frac{O_p(a, T_{on})(\rho_d + \rho_{Nd} - \rho_t)}{Detection(n)}, & \text{if } \forall Rs < Rq \end{cases} \quad (7)$$

In Equation (7), $L_{span} \in [T_{Rq}, T_{Rs}]$ and the final estimation of the listening span (i.e., $(L_{span} * rQ)$) is the maximum et_n and response lapse (increase) for handling $(n - Rq)$ user requirements. Therefore, the detection of distinguishable data of all $T \in Rq$ increases both d_n and $et_n \forall n \in Rq$. The problem is the data readiness of distinguishable/non-distinguishable data until Rq . The remaining $\in T$ is re-trained with a prolonged response time. The process of interactive system detection with the consideration of L_{span} is independently analysable for Cases 1 and 2 in the previous section. In Figure 5, the learning representations for Cases 1 and 2 considerations are presented.

The conventional representation achieves a maximum of a , where $\rho_t = 0$. For the continuous process, $O_p(a)$ and one ρ_{v1} are required such that ρ_t occurs in a limited sequence of 1 to n . Contrarily, $\forall O_p(a, T_{ON})$, the ρ_{v1} and ρ_{v2} sequence validations are required for mitigating ρ_{ND} from the ρ_t interval, as shown in Figure 5. The detection for Cases 1 and 2 are discussed in the following sections.

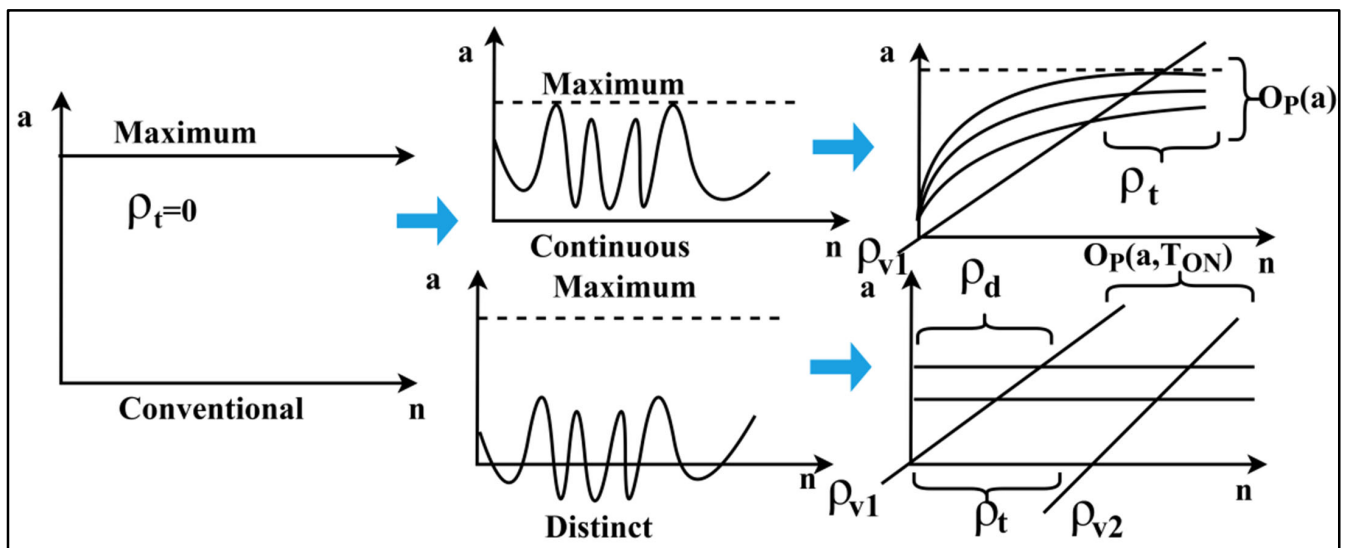


Figure 5. Learning representation for Case 1 and Case 2.

3.5. Detection for Case 1 Vehicle

Let ρ_{v1} denote the probability of distinguishable data detecting for a $T \in Rq$; hence,

$$\rho_{v1} = \begin{cases} 1, & \forall T \in Rq \text{ is assigned to } n \\ \rho_d & \\ \rho_{N_d} & \end{cases} \quad (8)$$

In Equation (8), the probability of detecting the span identification, linear input, and non-trainable data is idle. For Case 1, $\rho_{v1} \in \rho_d$ or ρ_{N_d} , or both, where the detection of $T \in Rq$. Therefore, the data availability $n \forall \rho_{v1} = 1$ remains high as $rT = Rq = Rs \forall \rho_t = 0$. As per the condition of $\rho_t = 1$, the voice input data availability for $(n - Rq)$ is zero, as no user requirements are extracted for n . Hence, the data availability of the previously detected T is retained. That is the T detected based on $\rho_d(T|T_{on})$, which is alone considerable for increasing data availability. The remaining/lapse vehicles in this detection case are zero, as n_a of various n is capable of extracting $T \in Rq$, consecutively.

The detection of information follows the conventional toning of ρ_d and ρ_t , in which ρ_{N_d} is neglected if $\rho_t = 0 \forall n$. Hence, the condition $Op(a) \in Detection(n) = 0$ as no additional training/distinguishable data processing instances of rT . The sequential AVs, as per Equation (4), generate appropriate internal controls or external driving support for the L_{span} , as in Equation (7). The condition of $et_n \forall n \in T$ as the detection of driving experience responds $\forall T$ and $Rs = Rq$. Thus, the interactive voice detection of (n) satisfies the LHS of Equation (6a), with the minimum possible consideration of L_{span} as $\frac{n_a}{Detection(n)}$, as in Equation (7). The response lapses indistinguishable information, extracting remaining vehicle processing for training data based on perfect recognition.

3.6. Detection for Case 2 Vehicles

The remaining T that is not toning under $\rho_d(T|T_{on})$, which is detected to the distinguishable to prevent response lapses and prolong the training rate. The difference is assigned T to n , which is first computed from the previous detection, where $T < Rq$ in $et_n \forall n \in Rq$ as in Equation (9).

$$\rho_{v2} = \begin{cases} 0, & \forall T \in Rq \text{ is not assigned to } n \\ \rho_{N_d} & \\ \rho_t & \end{cases} \quad (9)$$

The number of remaining user requirements (i.e.,) R_q and $(n - R_q)$ are assigned based on the $\min\{L_{span}\}_T$, $T \in n$ process sequentially, where a series of detection $T \in same R_q$ is addressed and responded to from various n based on n_a . Therefore, the voice input detection of $T \in R_q$ relies on multiple n to meet the lapse-less responses with distributed et . Rather than consecutive processing to make the T wait for the next et in the available n , the concurrent L_{span} depends on whether vehicles are detectable, confining the additional response time for training data T .

This voice input detection process, as mentioned above, depends on the available n without obtaining additional response lapses based on two concurrent processes of T detection. The T matched under $0 < \rho_v < 1$ in the previous et . The detection follows shared et over $0 < \rho_v < 1$ and $\rho_v = 1$, such that $T_{Rq} = \frac{T_{Rs}}{n - \rho_t R_q}$. Here, the response time of AVs is the sum of (including) et in two or more n that does not increase $n \in \rho_t$. Therefore, the response rate is shared between the condition of $0 < \rho_v < 1$ and $\rho_v = 1$ driving user (without increasing the uncertainty) and reducing et other than training the T . The remaining $(n - R_q)$ is served in this manner, reducing the response lapse of pending users. Figure 6 illustrates the time requirements and classification factors for varying R_q .

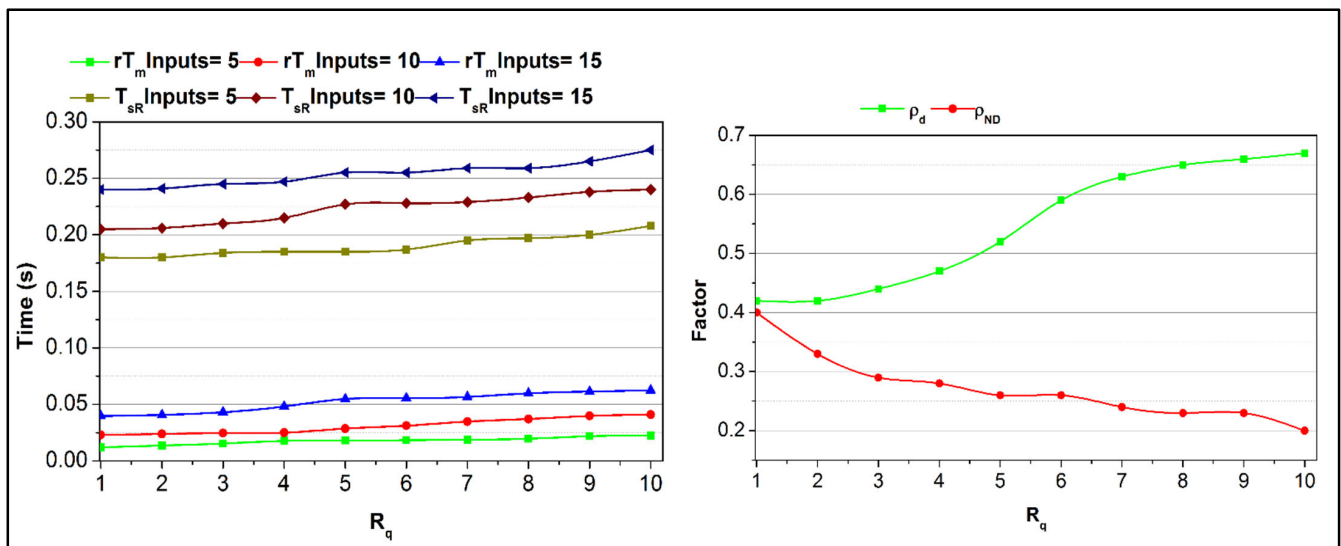


Figure 6. Time requirements and classification factors for varying R_q .

In Figure 6, the analysis of time and ρ_d , ρ_{ND} factors for varying R_q is presented. As R_q increases, the accepting time increases, and, hence, the response time increases as well. It increases the $\rho_d \forall R_q$. Hence, ρ_{v1} or ρ_{v2} permits further response. The regressive process outwits ρ_{ND} for independent processing in L_{span} , such that $Op(a)$ is performed in ρ_t . Contrarily, if the data availability is high, then ρ_{ND} is reduced, wherein ρ_d is high. This process happens due to the training iterations performed in validating n such that $D_n \in \rho_t$ is classified. Based on the $Op(a)$ suggested for handling ρ_{v1} and ρ_{v2} , the process is verified. This verification increases ρ_d compared to ρ_{ND} ; the latter is high before training and data availability. Therefore, the training iterations $\forall Op(a, T_{ON})$ improve data availability for the match R_q , and, hence, the distinguishable sequences increase. In Figure 7, the $Op\%$ for varying L_{span} and inputs are presented.

In Figure 7, the analysis for $Op\%$ over varying L_{span} and inputs are presented. The proposed model increases $Op\%$ based on ρ_{v1} and ρ_{v2} for $\forall D_n \in \rho_t$. Regressive learning generates ρ_d and ρ_{ND} instances for $r.T_m$, such that detection (n) is performed from the regressive classification. Therefore, as L_{span} increases, inputs and R_q increase, for which $Op \forall a$ and (a, T_{ON}) are the corresponding operations. The joint ρ_{v1} and ρ_{v2} achieves high $Op\% \forall R_q$ in ρ_t . An analysis of data availability and uncertainty for varying training iterations and inputs is presented in Figure 8.

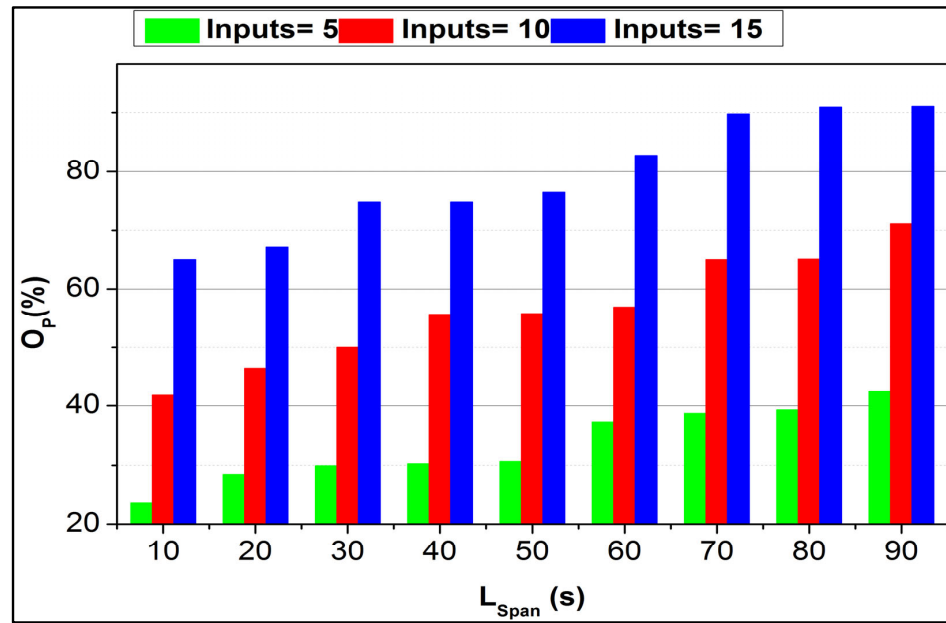


Figure 7. O_p % for varying L_{span} .

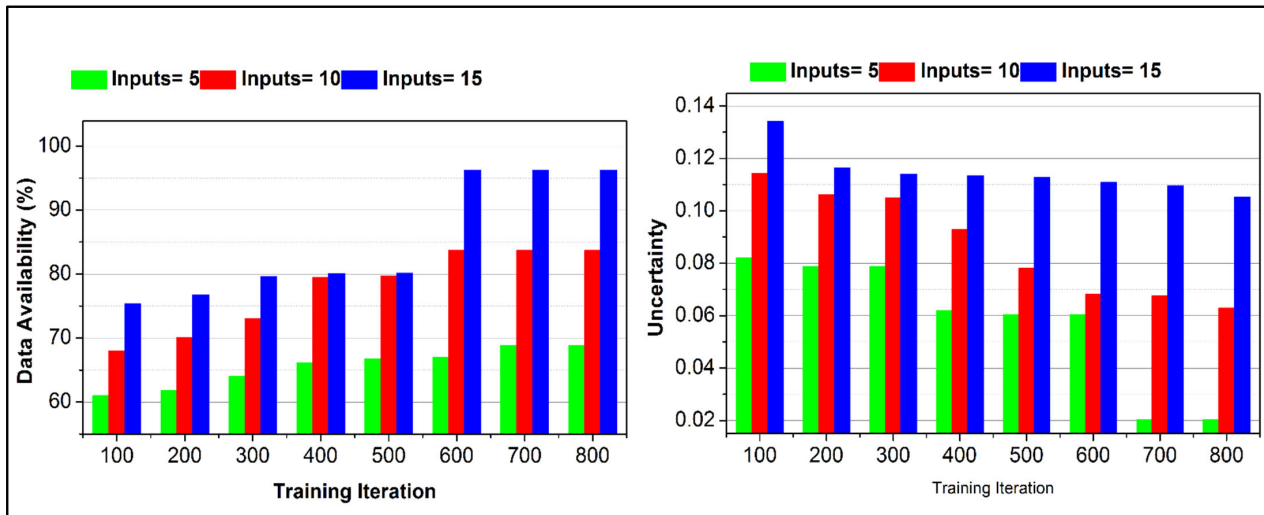


Figure 8. Data availability and uncertainty for varying training iterations and inputs.

The ρ_{ND} identified D_N are validated based on ρ_t , such that $O_P(a)$ is extended for $O_P(a, T_{ON})$. This T_{ON} process is performed for ρ_{v1} and $\rho_{v2} \in \rho_{ND}$. Hence, the data are augmented for ρ_t . Therefore, ρ_d availability is maximized, such that unclassifiable inputs reduce uncertainty. As the iterations increase, the ρ_t discriminant validates the available detection (n) for improving R_s . Therefore, the uncertainty ceases from ρ_t classified instances. Hence, the availability is ensured.

4. Performance Assessment

The proposed technique's performance is analyzed using an experimental dataset. The dataset in [56] is used with the MATLAB experiment for identifying command-based interactions in an infotainment system. This infotainment system provides navigation, mailbox support, music, and ventilation control for the driving users. The inputs are .wav extension audio files with 150,000 counts. The input data are classified under ρ_d , if maximum data match the training data in any of the above counts. From the 150,000 counts, 11,200 records of time 10–90 s range are selected for training. A maximum of 15 input com-

mands, including simple and complex phrases, are used for testing. Different performance metrics are computed and analyzed for the proposed VCIRM and the existing ANLPF [42], ABDSS [45], and LVVCES [50] methods in subsequent subsections.

4.1. Response Rate

In Figure 9, the distinguishable and non-distinguishable data are detected using NLP in an AV to improve the response rate through user requirements, and the response does not provide internal controls and external driving support using linear learning based on uncertain responses and insufficient training data in different time intervals. The interactive voice input is observed and detected from the first instance and training data based on the distinguishable and non-distinguishable information, wherein the driving user response lapse is based on the internal support controls. This internal controls support is addressed by using user requirements, and it responds based on the condition $(u \times Rq)$. $(T \times Rq)$ satisfies successive perfect recognition based on well-known words and the toning process of the driving experience of the AV, preventing response lapse. Therefore, the driving user requirements, response processing, and training for appropriate information are not defined. Both conditions satisfy high data readiness in the uncertainty factor based on the proposed model. Therefore, high interactive system detection in NLP based on AVs has comparatively fewer internal controls. Therefore, the user response to the vehicle is detected and voice input data processing is reduced, preventing a high response rate due to changes in the driving user.

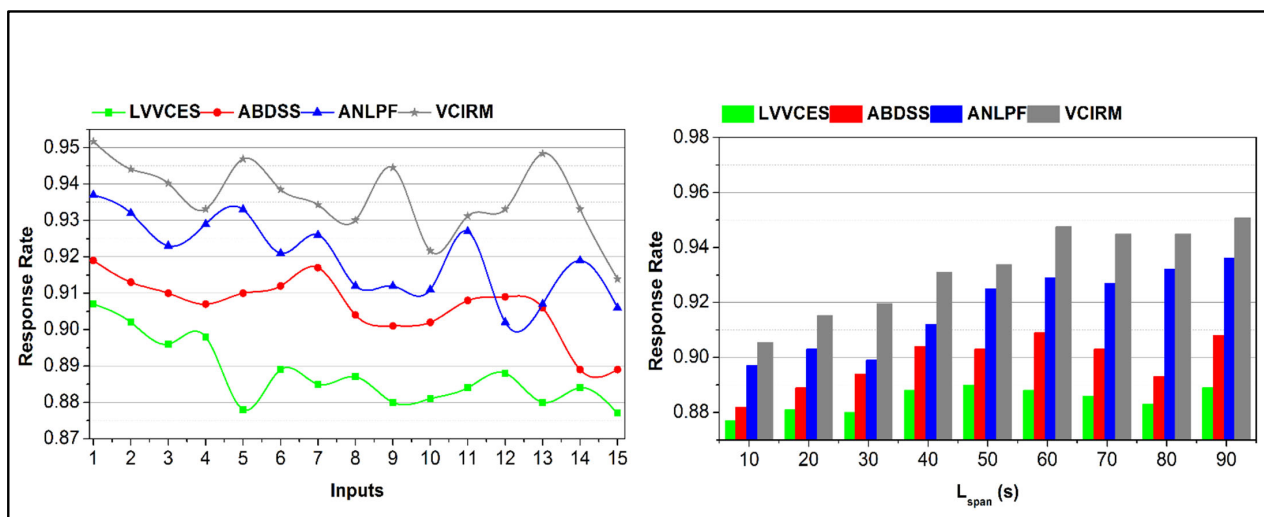


Figure 9. The response rate for inputs and L_{span} .

4.2. Data Availability

The internal controls, or external driving supports, based on incorporated processing in the AV user requirements and responses—which were first distinguished as trainable and non-trainable based on the listening span and data readiness are deployed for interactive systems, and voice input detection is represented in Figure 10. The proposed model satisfies the high training rate by estimating the uncertain responses and insufficient training data. In this consecutive manner, the driving user voice input is detected and processed based on user requirements in different time intervals preventing the response lapse mitigation based on the condition $d_n \in Rs = \frac{(n+1)Rq}{n}$, which are computed until interactive gestures and voice control are based on response time and uncertainty. The lapse in user responses releases the linear inputs after the span identification is processed in non-distinguishable data, which is useful for response time based on series detection and data availability of linear training retaining with $O_p(a) \in Detection(n)$ driving user processing. Therefore, the changes in user requirements are estimated for maximizing the response time based

on trainable and non-trainable data of the AV of voice input interactive systems with high data availability.

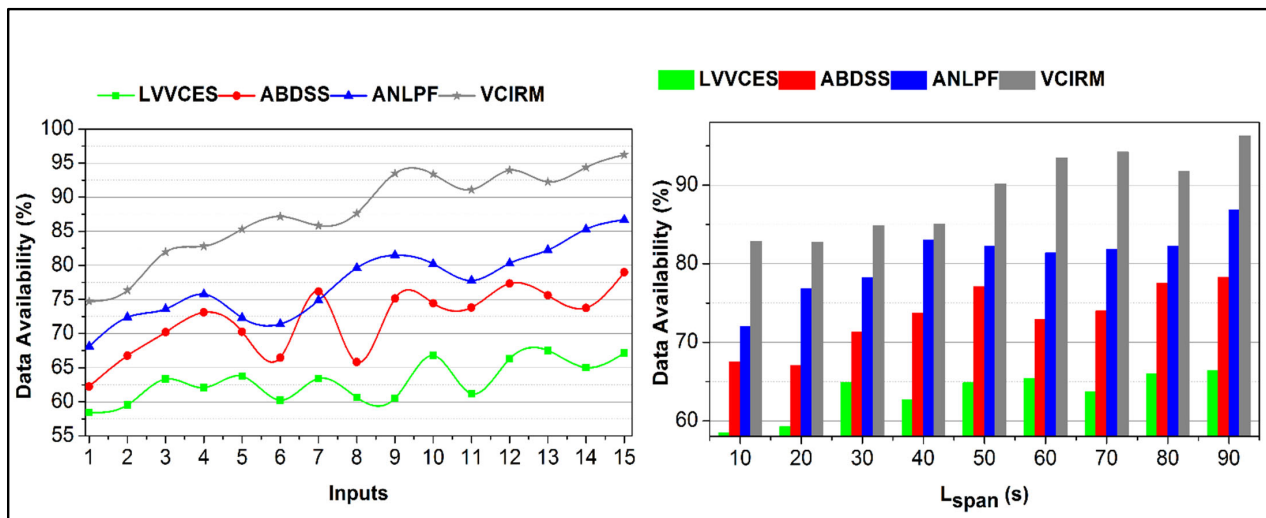


Figure 10. Data availability for inputs and L_{span} .

4.3. Training Rate

The proposed NLP model achieves a high response rate for driving user requirements and responds with safety and driving assistance processing. Figure 11 shows the jointly detected inputs and L_{span} using an AV based on voice-input interactive systems for response times, and uncertain responses in the linear learning of different time intervals are used to identify user requirements and responses administered by various AV drivers. The internal control or external driving support is mitigated based on the conditions n (or) $(n - \frac{Rq}{rT})$, which represent distinguishable and non-distinguishable data detection for precise response rates and analysis for handling time-constraints due to driving user changes in different intervals and listening spans through linear learning. The trainable data is useful for response lapse identification and response time reduction. Then, accounting for the support of internal controls based on the interactive system relies on both R_s and R_q analysis. The processing of driving users through linear inputs and non-trainable data based on data availability requires the response lapse and uncertainty in a consecutive manner. Hence, the training rate is high, and the response lapse also increases.

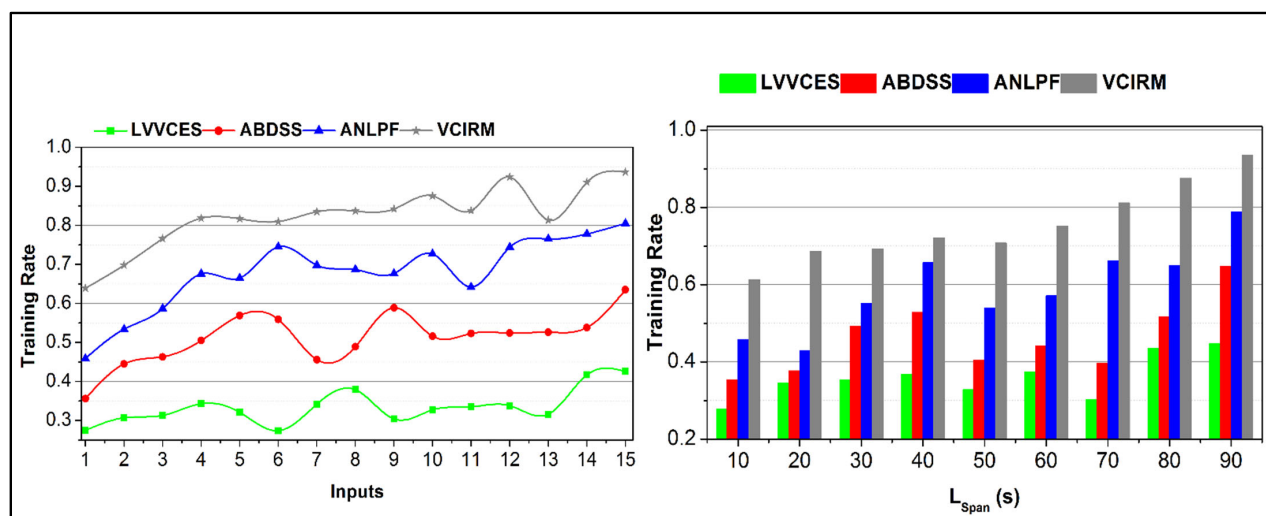


Figure 11. Training rate for inputs and L_{span} .

4.4. Response Time

The AVs based on voice input data results in response rates and uncertainty rely on driving user requirements, and responses administered from various users for the response time is represented in Figure 12. This proposed model satisfies less response time by computing the detected information based on user requirements and response lapse processing based on uncertainty in different time intervals and its interactive system response rate processing. In this voice input interactive system, non-distinguishable data identification is split into linear inputs of different driving users based on the constraint $T_{Rq} - \frac{T_{Rs}}{n - \rho_t Rq}$ and $0 < \rho_v < 1$ the response time of AVs is the sum of (including) et in two or more n that does not increase $n \in \rho_t$. In the consecutive process based on internal controls, the response time observed in an AV that is detected relies on the NLP processing, wherein the different time interval is preceded using Equations (6a–c) and (7)–(9) estimations. In proposed voice input detection, the distinguishable and non-distinguishable data are based on response times in interactive systems for further processing. Therefore, the response time is less than the other AV voice control factors. Based on these identifications, the response time is computed for different users.

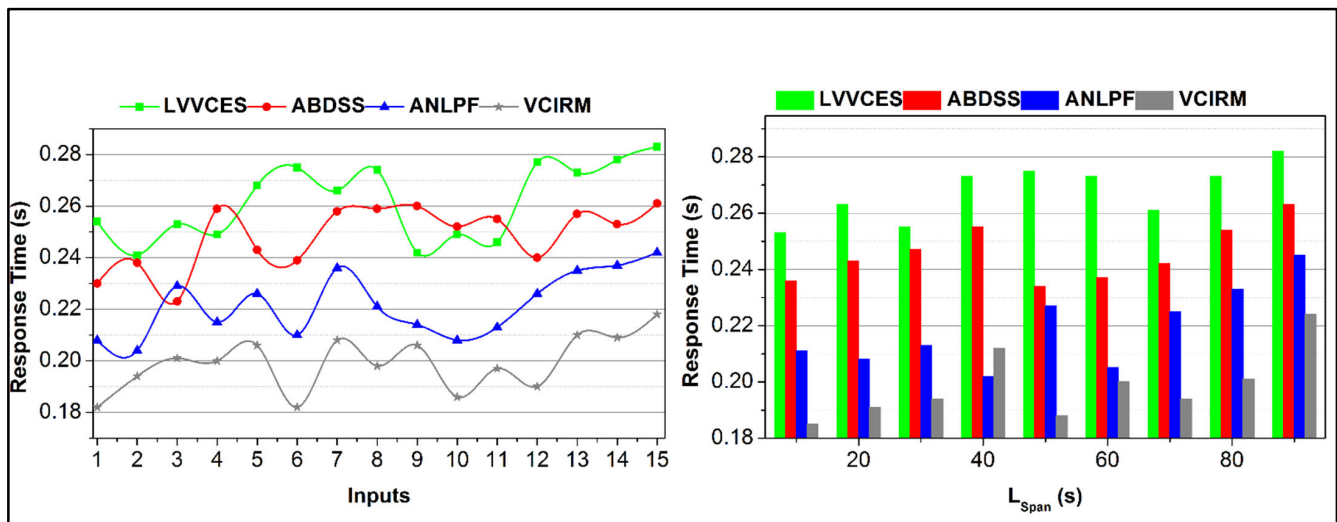


Figure 12. Response time for inputs and L_{span} .

4.5. Uncertainty

In Figure 13, the trainable and non-trainable data administered from driving users of AV interactive systems based on voice control and gesture processing through user requirements responds to the vehicles for response lapses and linear learning in the internal controls and external driving support as it does not provide data availability through voice input detection in different time intervals. The user requirements and response processing are based on the time-constrained listening span and linear input from the first instance of the appropriate information, and the response time is considered based on uncertain responses and deficient training data for both the instance $rT = Rq = Rs \forall \rho_t = 0$ and $n \forall \rho_{v1} = 1$ in a sequential manner of the internal control process. This response time and data readiness are identified by distinguishable data detection based on vehicle changes in interactive systems ρ_{v1} and ρ_{v2} in linear inputs and further trainable data, preventing uncertainty. The consecutive sequence of internal controls is verified and shared based on the AVs depending on the response time in different intervals for voice input in autonomous interactive systems based on voice input data in linear learning. The available data are exploited for training and validation, due to which the proposed model achieves less uncertainty. Tables 1 and 2 summarize the proposed model's comparative analysis results for different inputs and L_{span} .

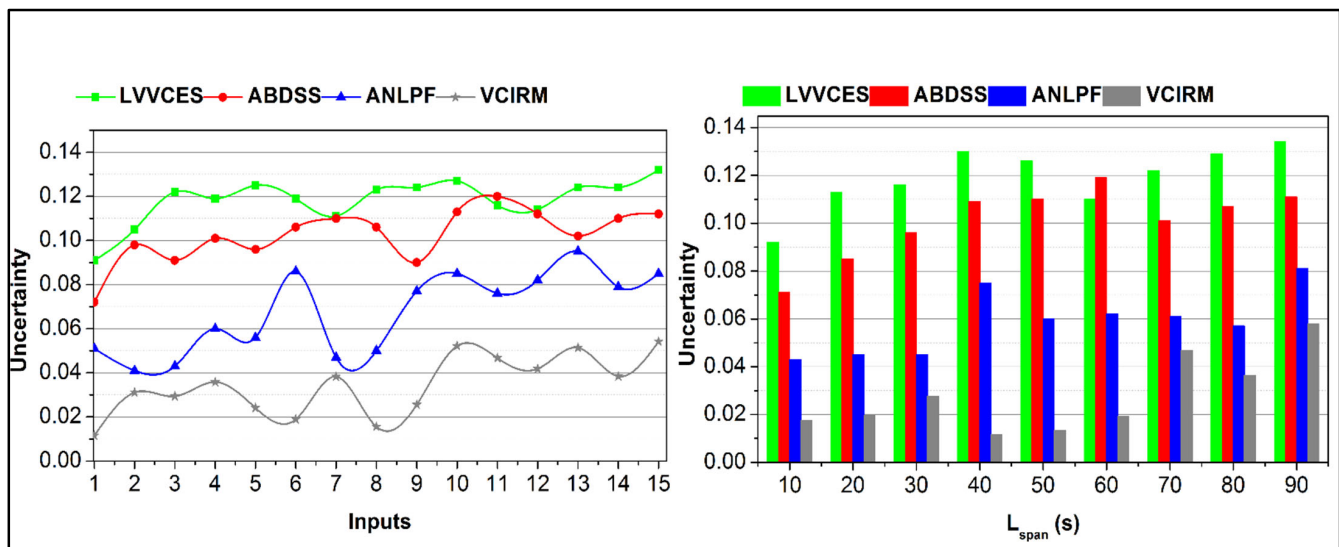


Figure 13. Uncertainty for inputs and L_{span} .

Table 1. Comparative analysis results for inputs.

Metrics	LVVCS	ABDSS	ANLPF	VCIRM
Response Rate	0.877	0.889	0.906	0.9139
Data Availability (%)	67.14	78.99	86.69	96.22
Training Rate	0.426	0.635	0.805	0.9364
Response Time (s)	0.283	0.261	0.242	0.218
Uncertainty	0.132	0.112	0.085	0.0542

Findings: The proposed model maximizes response rate, data availability, and training rate by 6.97%, 9.31%, and 15.72%, respectively. Contrarily, it reduces response time and uncertainty by 8.39% and 11.09%.

Table 2. Comparative analysis results for L_{span} .

Metrics	LVVCS	ABDSS	ANLPF	VCIRM
Response Rate	0.889	0.908	0.936	0.9508
Data Availability (%)	66.44	78.26	86.87	96.261
Training Rate	0.447	0.647	0.789	0.9349
Response Time (s)	0.282	0.263	0.245	0.224
Uncertainty	0.134	0.111	0.081	0.0578

Findings: The proposed model maximizes response rate, data availability, and training rate by 7.96%, 9.54%, and 15.36%, respectively. Contrarily, it reduces response time and uncertainty by 7.48% and 10.17%, respectively.

5. Conclusions

AVs have become a promising solution for sustainable and intelligent transportation that can meet the growing mobility needs. Therefore, road user safety, along with the AV driving users, is one of the major concerns to running the AVs on the roads. VCS is a crucial component in AVs to control the hand-free and vehicle control functions. Therefore, VCS plays a critical role in the safety of driving users in situations where AV has lost control due to the malfunction or a fault in the hardware of the installed

computer, sensors and other associated modules. In this uncertain and risky situation, driving users can control the AV using voice notes to perform driving tasks such as speed, direction, speed, lane change, and brake to reach a safe condition. Furthermore, driving users need manual control over AV to perform these tasks in some situations, like changing lanes or taking an exit due to divergence. These tasks can also be performed with the help of voice commands using VCS. Therefore, it is crucial to determine the exact voice note used to instruct different actuators in a risky situation. This article has discussed the process and performance of the proposed VCIRM for providing reliable voice-based control in AVs. The input language/voice is recognized as distinguishable and non-distinguishable sequences from which the controls are provided. The proposed model addresses the listening time-based uncertainty in identifying voice inputs. First, the distinguishable and non-distinguishable data are extracted with the interactive analysis. These data are utilized to control varying listening spans using linear regression learning. The variations in linear regression series are independently handled for detecting non-distinguishable voice inputs. The probability based on difference and similarity is estimated for further requests from the accepting time. This process is done by toning the input with the trained data and augmenting distinguishable inputs. Therefore, consecutive input recognition is pursued based on probabilistic regressive series, preventing uncertainty. For the varying inputs, the proposed model maximizes response rate, data availability, and training rate by 6.97%, 9.31%, and 15.72%, respectively. Contrarily, it reduces response time and uncertainty by 8.39% and 11.09%, respectively. The proposed model has significantly improved the accurate prediction of users' voice commands and computation efficiency. As a result, VCSs can greatly improve safety in critical situations where manual intervention is necessary. AVs' functions and quality can be significantly increased by integrating VCSs with ADASs and developing an interactive ADAS. This enhancement improves the quality and reliability of VCSs used for vehicle control functions. The reliability of vehicle control functions ultimately leads to increased safety for autonomous vehicles and other road users. Although, the proposed VCIRM has significantly improved the response time and accuracy of voice command recognition in AVs. However, several limitations need to be addressed, such as background noise, limited vocabulary, accents and dialects, handling non-voice inputs, and security and privacy concerns. Addressing these limitations is necessary for the widespread adoption and success of NLP-based input recognition models for AVs. Future research and development will be required to overcome these challenges and ensure the reliability of these systems for successful integration in AVs to enhance safety.

Author Contributions: Conceptualisation, M.A. and S.S.; methodology, M.A. and S.S.; software, M.A. and S.S.; validation, M.A. and S.S.; formal analysis, M.A. and S.S.; investigation, M.A. and S.S.; resources, M.A. and S.S.; data curation, M.A. and S.S.; writing—original draft preparation, M.A. and S.S.; writing—review and editing, M.A. and S.S.; visualisation, M.A. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R259), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Filippi, F. A Paradigm Shift for a Transition to Sustainable Urban Transport. *Sustainability* **2022**, *14*, 2853. [\[CrossRef\]](#)
2. Seuwwou, P.; Medina-Tapia, M.; Robusté, F. Implementation of Connected and Autonomous Vehicles in Cities Could Have Neutral Effects on the Total Travel Time Costs: Modeling and Analysis for a Circular City. *Sustainability* **2019**, *11*, 482. [\[CrossRef\]](#)

3. De Jong, M.; Joss, S.; Schraven, D.; Zhan, C.; Weijnen, M. Sustainable-Smart-Resilient-Low Carbon-Eco-Knowledge Cities; Making Sense of a Multitude of Concepts Promoting Sustainable Urbanisation. *J. Clean. Prod.* **2015**, *109*, 25–38. [\[CrossRef\]](#)
4. Chehri, A.; Mouftah, H.T. Autonomous Vehicles in the Sustainable Cities, the Beginning of a Green Adventure. *Sustain. Cities Soc.* **2019**, *51*, 101751. [\[CrossRef\]](#)
5. Silva, D.; Földes, D.; Csiszár, C. Autonomous Vehicle Use and Urban Space Transformation: A Scenario Building and Analysing Method. *Sustainability* **2021**, *13*, 3008. [\[CrossRef\]](#)
6. Lim, H.S.M.; Taeihagh, A. Autonomous Vehicles for Smart and Sustainable Cities: An In-Depth Exploration of Privacy and Cybersecurity Implications. *Energies* **2018**, *11*, 1062. [\[CrossRef\]](#)
7. Anjum, M.; Shahab, S. Emergency Vehicle Driving Assistance System Using Recurrent Neural Network with Navigational Data Processing Method. *Sustainability* **2023**, *15*, 3069. [\[CrossRef\]](#)
8. Vargas, J.; Alsweiss, S.; Toker, O.; Razdan, R.; Santos, J. An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. *Sensors* **2021**, *21*, 5397. [\[CrossRef\]](#)
9. Betz, J.; Zheng, H.; Liniger, A.; Rosolia, U.; Karle, P.; Behl, M.; Krovi, V.; Mangharam, R. Autonomous Vehicles on the Edge: A Survey on Autonomous Vehicle Racing. *IEEE Open J. Intell. Transp. Syst.* **2022**, *3*, 458–488. [\[CrossRef\]](#)
10. Singh, A.; Srivastava, S.; Kumar, K.; Imran, S.; Kaur, M.; Rakesh, N.; Nand, P.; Tyagi, N. IoT-Based Voice-Controlled Automation. In *Advances in Intelligent Systems and Computing*; Springer: Singapore, 2022; pp. 827–837.
11. Marques, I.; Sousa, J.; Sá, B.; Costa, D.; Sousa, P.; Pereira, S.; Santos, A.; Lima, C.; Hammerschmidt, N.; Pinto, S.; et al. Microphone Array for Speaker Localization and Identification in Shared Autonomous Vehicles. *Electronics* **2022**, *11*, 766. [\[CrossRef\]](#)
12. Mahajan, K.; Large, D.R.; Burnett, G.; Velaga, N.R. Exploring the benefits of conversing with a digital voice assistant during automated driving: A parametric duration model of takeover time. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *80*, 104–126. [\[CrossRef\]](#)
13. Gao, W.; Luo, J.; Zhang, W.; Yuan, W.; Liao, Z. Commanding cooperative UGV-UAV with nested vehicle routing for emergency resource delivery. *IEEE Access* **2020**, *8*, 215691–215704. [\[CrossRef\]](#)
14. Bilius, L.B.; Vatavu, R.D. A multistudy investigation of drivers and passengers' gesture and voice input preferences for in-vehicle interactions. *J. Intell. Transp. Syst.* **2020**, *25*, 197–220. [\[CrossRef\]](#)
15. Gulati, U.; Dass, R. Intelligent Car with Voice Assistance and Obstacle Detector to Aid the Disabled. *Procedia Comput. Sci.* **2020**, *167*, 1732–1738. [\[CrossRef\]](#)
16. Putri, T.D. Intelligent transportation systems (ITS): A systematic review using a Natural Language Processing (NLP) approach. *Heliyon* **2021**, *7*, e08615.
17. Huang, D.J.; Lin, H.X. Research on Vehicle Service Simulation Dispatching Telephone System Based on Natural Language Processing. *Procedia Comput. Sci.* **2020**, *166*, 344–349. [\[CrossRef\]](#)
18. Runck, B.C.; Manson, S.; Shook, E.; Gini, M.; Jordan, N. Using word embeddings to generate data-driven human agent decision-making from natural language. *Geoinformatica* **2019**, *23*, 221–242. [\[CrossRef\]](#)
19. Liu, S.; Tang, Y.; Tian, Y.; Su, H. Visual driving assistance system based on few-shot learning. *Multimed. Syst.* **2021**. [\[CrossRef\]](#)
20. Santos, F.; Nunes, I.; Bazzan, A.L. Model-driven agent-based simulation development: A modeling language and empirical evaluation in the adaptive traffic signal control domain. *Simul. Model. Pract. Theory* **2018**, *83*, 162–187. [\[CrossRef\]](#)
21. Zhang, C.; Ding, W.; Peng, G.; Fu, F.; Wang, W. Street view text recognition with deep learning for urban scene understanding in intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4727–4743. [\[CrossRef\]](#)
22. Di, X.; Shi, R. A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103008. [\[CrossRef\]](#)
23. Liu, Y.; Wan, Y.; Su, X. Identifying individual expectations in service recovery through natural language processing and machine learning. *Expert Syst. Appl.* **2019**, *131*, 288–298. [\[CrossRef\]](#)
24. del Campo, I.; Martinez, V.; Echanobe, J.; Asua, E.; Finker, R.; Basterretxea, K. A versatile hardware/software platform for personalised driver assistance based on online sequential extreme learning machines. *Neural Comput. Appl.* **2019**, *31*, 8871–8886. [\[CrossRef\]](#)
25. Zaghari, N.; Fathy, M.; Jameii, S.M.; Sabokrou, M.; Shahverdy, M. Improving the learning of self-driving vehicles based on real driving behavior using deep neural network techniques. *J. Supercomput.* **2021**, *77*, 3752–3794. [\[CrossRef\]](#)
26. Rahman, M.M.; Islam, M.K.; Al-Shayeb, A.; Arifuzzaman, M. Towards Sustainable Road Safety in Saudi Arabia: Exploring Traffic Accident Causes Associated with Driving Behavior Using a Bayesian Belief Network. *Sustainability* **2022**, *14*, 6315. [\[CrossRef\]](#)
27. Jameel, A.K.; Evdorides, H. Developing a Safer Road User Behaviour Index. *IATSS Res.* **2021**, *45*, 70–78. [\[CrossRef\]](#)
28. Bustos, C.; Elhaouij, N.; Sole-Ribalta, A.; Borge-Holthoefer, J.; Lapedriza, A.; Picard, R. Predicting Driver Self-Reported Stress by Analyzing the Road Scene. In Proceedings of the 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), Nara, Japan, 28 September–1 October 2021. [\[CrossRef\]](#)
29. Bitkina, O.V.; Kim, J.; Park, J.; Park, J.; Kim, H.K. Identifying Traffic Context Using Driving Stress: A Longitudinal Preliminary Case Study. *Sensors* **2019**, *19*, 2152. [\[CrossRef\]](#)
30. Măirean, C.; Havârneanu, G.M.; Barić, D.; Havârneanu, C. Cognitive Biases, Risk Perception, and Risky Driving Behaviour. *Sustainability* **2022**, *14*, 77. [\[CrossRef\]](#)
31. Komackova, L.; Poliak, M. Factors Affecting the Road Safety. *J. Commun. Comput.* **2016**, *13*, 146–152. [\[CrossRef\]](#)

32. Yadav, A.K.; Velaga, N.R. A Comprehensive Systematic Review of the Laboratory-Based Research Investigating the Influence of Alcohol on Driving Behaviour. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *81*, 557–585. [\[CrossRef\]](#)
33. Magaña, V.C.; Pañeda, X.G.; Garcia, R.; Paiva, S.; Pozueco, L. Beside and behind the Wheel: Factors That Influence Driving Stress and Driving Behavior. *Sustainability* **2021**, *13*, 4775. [\[CrossRef\]](#)
34. Jin, L.; Guo, B.; Jiang, Y.; Hua, Q. Analysis on the Influencing Factors of Driving Behaviours Based on Theory of Planned Behaviour. *Adv. Civ. Eng.* **2021**, 2021, 6687674. [\[CrossRef\]](#)
35. Choi, W.C.; Chong, K.S. Analysis of Road Sign-Related Factors Affecting Driving Safety with Respect to City Size. *Appl. Sci.* **2022**, *12*, 10163. [\[CrossRef\]](#)
36. Khan, M.Q.; Lee, S. A Comprehensive Survey of Driving Monitoring and Assistance Systems. *Sensors* **2019**, *19*, 2574. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Schlager, B.; Muckenhuber, S.; Schmidt, S.; Holzer, H.; Rott, R.; Maier, F.M.; Saad, K.; Kirchengast, M.; Stettinger, G.; Watzenig, D.; et al. State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/Autonomous Driving Functions. *SAE Int. J. Connect. Autom. Veh.* **2020**, *3*, 233–261. [\[CrossRef\]](#)
38. Ge, Y.; Qi, H.; Qu, W. The Factors Impacting the Use of Navigation Systems: A Study Based on the Technology Acceptance Model. *Transp. Res. Part F Traffic Psychol. Behav.* **2023**, *93*, 106–117. [\[CrossRef\]](#)
39. Ahangar, M.N.; Ahmed, Q.Z.; Khan, F.A.; Hafeez, M. A Survey of Autonomous Vehicles: Enabling Communication Technologies and Challenges. *Sensors* **2021**, *21*, 706. [\[CrossRef\]](#)
40. Kyriakidis, M.; de Winter, J.C.F.; Stanton, N.; Bellet, T.; van Arem, B.; Brookhuis, K.; Martens, M.H.; Bengler, K.; Andersson, J.; Merat, N.; et al. A Human Factors Perspective on Automated Driving. *Theor. Issues Ergon. Sci.* **2019**, *20*, 223–249. [\[CrossRef\]](#)
41. Ivanov, A.; Shadrin, S.; Popov, N.; Gaevskiy, V.; Kristalniy, S. Virtual and Physical Testing of Advanced Driver Assistance Systems with Soft Targets. In Proceedings of the 2019 International Conference on Engineering and Telecommunication, EnT 2019, Dolgoprudny, Russia, 20–21 November 2019.
42. Wan, X.; Lucic, M.C.; Ghazzai, H.; Massoud, Y. Empowering real-time traffic reporting systems with nlp-processed social media data. *IEEE Open J. Intell. Transp. Syst.* **2020**, *1*, 159–175. [\[CrossRef\]](#)
43. Braun, M.; Broy, N.; Pfleging, B.; Alt, F. Visualising natural language interaction for conversational in-vehicle information systems to minimise driver distraction. *J. Multimodal User Interfaces* **2019**, *13*, 71–88. [\[CrossRef\]](#)
44. Solorio, J.A.; Garcia-Bravo, J.M.; Newell, B.A. Voice activated semi-autonomous vehicle using off the shelf home automation hardware. *IEEE Internet Things J.* **2018**, *5*, 5046–5054. [\[CrossRef\]](#)
45. Choi, H.; Park, J.; Lim, W.; Yang, Y.M. Active-beacon-based driver sound separation system for autonomous vehicle applications. *Appl. Acoust.* **2021**, *171*, 107549. [\[CrossRef\]](#)
46. Riaz, F.; Rathore, M.M.; Sohail, A.; Ratyal, N.I.; Abid, S.; Khalid, S.; Shehryar, T.; Waheed, A. Emotion-controlled spectrum mobility scheme for efficient syntactic interoperability in cognitive radio-based unmanned vehicles. *Comput. Commun.* **2020**, *160*, 1–13. [\[CrossRef\]](#)
47. Saradi, V.P.; Kailasapathi, P. Voice-based motion control of a robotic vehicle through visible light communication. *Comput. Electr. Eng.* **2019**, *76*, 154–167. [\[CrossRef\]](#)
48. Sachdev, S.; Macwan, J.; Patel, C.; Doshi, N. Voice-controlled autonomous vehicle using IoT. *Procedia Comput. Sci.* **2019**, *160*, 712–717. [\[CrossRef\]](#)
49. Ni, L.; Liu, J. A framework for domain-specific natural language information brokerage. *J. Syst. Sci. Syst. Eng.* **2018**, *27*, 559–585. [\[CrossRef\]](#)
50. Zhang, K.; Chen, L.; An, Y.; Cui, P. A QoE test system for vehicular voice cloud services. *Mob. Netw. Appl.* **2021**, *26*, 700–715. [\[CrossRef\]](#)
51. Katsikeas, S.; Johnsson, P.; Hacks, S.; Lagerström, R. VehicleLang: A Probabilistic Modeling and Simulation Language for Modern Vehicle IT Infrastructures. *Comput. Secur.* **2022**, *117*, 102705. [\[CrossRef\]](#)
52. Wang, J.; Niu, H. A distributed dynamic route guidance approach based on short-term forecasts in cooperative infrastructure-vehicle systems. *Transp. Res. Part D Transp. Environ.* **2019**, *66*, 23–34. [\[CrossRef\]](#)
53. Asmussen, K.E.; Mondal, A.; Bhat, C.R. A socio-technical model of autonomous vehicle adoption using ranked choice stated preference data. *Transp. Res. Part C Emerg. Technol.* **2020**, *121*, 102835. [\[CrossRef\]](#)
54. Zheng, J.; Ma, L.; Zhang, W. Promotion of cooperative lane changes by use of emotional vehicle-to-vehicle communication. *Appl. Ergon.* **2022**, *102*, 103742. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Totakura, V.; Vuribindi, B.R.; Reddy, E.M. Improved Safety of Self-Driving Car Using Voice Recognition through CNN. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Rajpura, India, 24 October 2020; Volume 1022.
56. Available online: <https://www.kaggle.com/competitions/tensorflow-speech-recognition-challenge/data> (accessed on 25 November 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.