The study [7] gave a description of pioneering approach regarding voice control of operative and functions which are technical in a real-world environment of smart home in which the control of voice within smart home is essential to grant systems of robust technology for establishing automation and for visualization of technology, software intended to recognize voice commands of an individual and a system which is robust for cancelling additive noise. However Independent Component Analysis (ICA) method which was utilized in this study has its limitations. These limitations include the inability of recovering source signals energy and inability to maintain source signal order. Thus, component output posses' distinct amplitude in relation to input signals, and when the method of ICA is adapted again, the component possesses a distinct order and signals of polarity.

The next section of the paper is voice activated systems challenges, this section outlines into detail the inability of automated inference of audio data in voice activated systems, the section of voice activated systems challenges also embodies the research question which this study sought to address. The third section is the research methodology, which outlines the specific procedures and methods employed to conduct this study. This section enables the reader to assess the soundness and accuracy of this study. The fourth section is the literature review and discussion, which offers a critical synthesis of existing knowledge regarding automated inference and command of audio data into voice activated systems and further provides an analysis of the studies explored in the literature review and also highlights the findings from this analysis. The last section is the conclusion and future work which summarises the findings of the study being pursued and addresses work to be carried out, derived from the current study.

2. Voice Activated Systems Challenges

Smart Homes collect annotated domestic environment audio recordings and store them in a database. However, this audio database is not used to personalize the smart systems that the user/inhabitant comes across in a different home. The challenge with voice activated systems is their inability to automatically inferring audio data. Voice activated systems rely on being operated manually by the user. Voice activated systems are pre-programmed to perform tasks based on computerised settings. For an example, if a user owns a smart home voice activated system and decides to replace it with a new one, they will need to manually input audio data into the new system. The new system will not automatically recognise the voice of its new user, extract audio data from the previous smart device database, and be able to configure itself in the same manner as the user of the previous smart device. This leads to the research question that this study aims to address: what machine learning techniques can be used to achieve automated inference of audio data in voice activated systems? Addressing this research question may result in seamless user mobility across voice-activated systems, which remains a challenge in the smart homes domain.

3. Research Methodology

A research methodology was adopted for this study to provide principled guidance and systematic plan of carrying out this study. A methodology is an embodiment of techniques, postulation, and regulations exploited by a discipline [8]. It is vitally important to follow a specific methodology in order to make it easy for the intended audience to rely on the results which have been obtained in this study. A methodology legitimises and dispenses sound scientific findings [9]. The adopted research methodology ensured that the study was carried out logically and provided guidance in conducting the study. It rendered a comprehensive plan and assisted in keeping researchers in line, facilitating the entire process, making if efficient and controllable, as explained in [10]. Thus, this section is critical for the overall soundness and accuracy of the whole study, as advanced in [11]. The adopted methodology is influenced by previous work [12] with the purpose of obtaining advanced outcomes to address the research question.

The research methodology consists of four subsections. The first subsection is the inclusion criteria, which provides details regarding the concepts of studies included to inform the methodology for conducting the literature review. The second subsection is literature identification, which involves specifying the databases used to obtain the studies included in this study, as well as outlining restrictions on the search date for included studies. The third subsection is screening for inclusion, which reflects the process undertaken to determine which studies are included. The fourth subsection is quality and eligibility assessment, which provides details on the criteria that influenced the inclusion of studies while considering the credibility of the studies included in this systematic literature review. Below, we provide a more detailed explanation of these subsections.

3.1. Inclusion Criterion

The criteria targeted articles that addressed the following two concepts, namely Voice Command and Machine Learning, and Automated Inference. Any articles that addressed these concepts in different contexts were excluded.

3.2. Literature Identification

The method used for literature identification involved a systemic search with the intension of acquiring relevant material. The databases used included IEEE Xplore, Science Direct and Springer Link. The search was carried out utilizing keywords ‘Voice Command and Machine Learning’, and ‘Automated Inference’. The preliminary relevance of each paper was determined based on its title. If the title appeared to discuss, at the minimum, audio data inference, we obtained the paper's full reference, overall purpose, methodology, and contributions. A publication date restriction was applied, allowing for papers published between 2019 and 2023, using the keyword Automated Inference.' For the keyword 'Voice Command and Machine Learning,' the search date was set from 2018 to 2023, ensuring that the review incorporated modern literature, considering data retrieval and integration in the computer age.

3.3. Screening for Inclusion

Numerous studies were retrieved through keyword searches. These retrieved studies were downloaded and subsequently subjected to inclusion and exclusion criteria. Papers relevant to audio data inference were selected for inclusion in this study, while those not relevant to the study were excluded. A search for the keyword 'Voice Command and Machine Learning' resulted in 104 pertinent studies. Out of these 104 studies, only 1 study was included, while 103 studies were excluded. A search for the keyword 'Automated Inference' resulted in 642 studies. After the first exclusion, 142 studies were removed, leaving 500. Among these studies, 1 study was included, while 499 studies were excluded. In the case of the Science Direct database, a search for the keyword 'Voice Command and Machine Learning' yielded 501 results. Out of these 501 studies, 3 studies were included, while 498 studies were excluded. In the Springer Link database, a search for the keyword 'Automated Inference' resulted in 34,903 hits. After the initial exclusion, 34,403 studies were removed. From the remaining 500 studies, 1 was included, and the other 499 were excluded.

3.4. Quality and Eligibility Assessment

A thorough perusal was conducted to determine the quality and suitability of studies. Articles from reputable publishers, namely IEEE Xplore, ScienceDirect and Springer Link with high research quality and peer-reviewed articles were subsequently selected for review.

4. Literature Review and Discussion

This section is divided into two subsections, namely the review of selected papers and analysis and discussion. This approach is used to ensure that each article is thoroughly reviewed and discussed in the context of the research question. Solutions, if any, are instantly identified and qualitied at length.

4.1. Review of Selected Papers

To solicit an understanding of existing knowledge regarding audio data inference in voice-activated systems and subsequently determine potent techniques, particularly machine learning techniques, literature was explored following the research methodology outlined above. To this end, literature was explored under two concepts: voice command and machine learning, and automated inference. The selected and reviewed articles are presented in Table 1 below. The table provides relevant details about each considered article, including the title, overall aim, methodology and contributions. This overview precedes the critique presented below the table.

Table 1: Existing research on audio data inference into voice-activated systems.

Concept	Paper	Overall Aim	Methodology	Contributions
Voice command and Machine Learning	Mao, J., Wang, C., Guo, Y., Xu, G., Cao, S., Zhang, X. and Bi, Z., 2022. A novel model for voice command fingerprinting using deep learning. Journal of Information Security and Applications, 65, p.103085	An investigation of deep learning in this study was pursued to advance the classification of the traffic which is encrypted in the systems of smart speaker not having to select features manually.	Features of simple input were fed in the model by the researchers of this study including interpacket time and information direction size and does not need artificial means of extraction of features.	The model attained an accuracy which surpass 93%. The model outmatches various Stat-of-Art models of voice command fingerprinting in the scenario of closed world. Experiments were also carried out by the researchers of this study on scenarios of open world and reflected the ability of the model to successfully differentiate voice commands which are of interest to an attacker. The researchers of this study evinced the ability of the model with regards to it being able to be applied to other respective domains which concern fingerprinting for instance website fingerprinting.
	Mondal, S. and Barman, A.D., 2022. Deep learning technique based real-time audio event detection experiment in a distributed system architecture. Computers and Electrical Engineering, 102, p.108252.	The overall aim of this paper is to investigate bandwidth efficient audio event detection technique and also noise robust	It is in the neural network scheme where the suggested denoising auto encoder – based completely joint scheme of deep neural network is realized. Extraction of audio features is conducted utilizing a Raspberry Pi edge device of computing which those to the local information server are fed for detection of audio events.	At F1- score achieved of 90.6% of signal noise ratio 10db and accuracy of 90% a reflection of advanced categorization likelihood of noisy events of audio.
	Filipe, L., Peres, R.S. and Tavares, R.M., 2021. Voice-activated smart home controller using machine learning. IEEE Access, 9, pp.66852-66863	The study pursued to present an implementation of an end-to-end smart home voice activated controller and design intended for devices	In order to achieve intelligent solution which is able to evolve and adapt hand in hand with the user techniques of online were exploited. The solution suggested also amalgamate the prospect of voice control, manage on application of mobile via an interface of a web. The features	• Architecture of smart home blueprint. • Implementation of end-to-end smart home controller and separate instructions. • Dataset of open source of the behaviour of the user

		which are intelligent and deployed in real environment. The validation was carried out in motorized blinds of experimental set up.	should be executed in a singular device light, small, portable, and compact, for prediction and adaption to the behaviours of users. Various modules must compose the controller in charge for various systems aspects. Denoted by figure 1 the architecture suggested for smart home controller is made up of specifically three modules. Adaptive Controller, Speech Recognition Platform, and Internet of Things Framework.	from the scenario of smart blind. • Contrast between approaches offline and online learning.
	Brenon, A., Portet, F. and Vacher, M., 2018. Arcades: A deep model for adaptive decision making in voice controlled smart home. Pervasive and Mobile Computing, 49, pp.92-110	Arcades are described in this study for extraction of context of representation of graphic of the system of home automation deep reinforcement learning is utilized and habitually updates its behaviour to that of the user.	It is greatly understood when it comes to Convolutional Neural Networks (CNN) that its requires large amount of data this prompted the researchers of this work to utilize CNN utilizing data generated artificially for feeding CNN with input which are raw for learning weights of CNN, execution of adapting decision model via reinforcement learning was performed, heterogenous information was presented and an approach of raw image was executed the decision model was based on deep learning.	Work submitted by this study provides the following contributions. • Users trivial interchange (rewards) is utilized by the system for adaption of the process of decision making. • For deciding the graphical representation of heterogenous data multi model can be translated. • The system can adjust its behaviour to that of the user and sensors which are not installed, or which are faulty. • Contextual data which is relevant can be learnt by the deep model. • Online learning can be carried out utilizing restricted history which is of the past. And absolutely no data pertaining to the future.

Automated Inference	Berns, F., Hüwel, J. and Beecks, C., 2022. Automated Model Inference for Gaussian Processes: An Overview of State-of-the-Art Methods and Algorithms. *SN Computer Science*, *3*(4), p.300	Recapitulation or outline of algorithms and methods of Stat-of-Art.	In the study analysis based on performance was executed on the following respective algorithm Compositional Kernel Search (CKS), Automatic Bayesian Covariance Discovery (ABCD), Scalable Kernel Composition (SKC), Large-Scale Automatic Retrieval (LARGEe), Concatenated Composite Covariance Search (3CS), and Lineage GPM Inference (LGI) possessing findings which are theoretical based the analysis of performance was conducted on the successive method's local approximations and global approximation and assessments were conducted on the successive algorithm LGI, ABCD, LARGe, CKS, 3CS and SKC.	The results of the study obtained denoted that approximated inference algorithms mainly locally approximating ones bring high level runtime performance preserving the calibre of those utilizing non-approximative Gaussian processes.
	Li, Y., Han, Z., Zhang, Q., Li, Z. and Tan, H., 2020, July. Automating cloud deployment for deep learning inference of real-time online services. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (pp. 1668-1677). IEEE.	Cloud deployment automation was pursued in this study for inference of online Deep Neural Network (DNN) in real time considering nominal cost subject to constraint.	The researchers of this study performed and implementation of a archetype system to their solution centred on the TensorFlow and carried out large-scale of experiments on Microsoft Azure.	The reflection made by the results obtained denote that the findings significantly outdo or surpass non-trivial baselines in relation to speed inference and also efficiency of costs.

Table 1: Concept, overall aim, methodology, and contributions of each study respectively.

From the Table 1 above, the researchers of each study had distinct aims, although their studies were based on Sound Event Detection. They successfully utilized various techniques to conduct their studies. These studies, as detailed in the table above, yielded positive results that significantly contribute to the field of Sound Event Detection. In the following section, an analysis of each study is performed, paying attention to the methodology used and highlighting the findings from the analysis.

4.2. Analysis and Discussion

The study seeks to obtain machine learning techniques which can be utilized to successfully achieve automated inference of audio data into voice activated systems. Techniques used in the study [13] do not only address automated inference but also have an added advantage which is of low latency in performance with low costs. Deep Neural Network was explored in the study utilizing algorithm Bayesian Optimization and Deep Reinforcement the results of the study showed that findings significantly surpass non-trivial baselines in relation to speed inference and efficiency of costs.

In the study [14] analysis based on performance was executed on the following respective algorithm Compositional Kernel Search (CKS), Automatic Bayesian Covariance Discovery (ABCD), Scalable Kernel Composition (SKC), Large-Scale Automatic Retrieval (LARGEe), Concatenated Composite Covariance Search (3CS), and Lineage GPM Inference (LGI) possessing findings which are theoretically based. The analysis of performance was conducted on the successive method's local approximations and global approximation and assessments were conducted on the successive algorithm LGI, ABCD, LARGe, CKS, 3CS and SKC. The results of the study obtained denoted that approximated inference algorithms mainly locally approximating ones bring high level runtime performance preserving the calibre of those utilizing non-approximative Gaussian processes.

The architecture suggested in the study [15] was after making good use of Machine Learning attributes like its capability to grasp and differentiate patterns that are not noticeable to humankind. Online learning methods were utilized for means of achieving a solution which is intelligent of being able to adapt and evolve with the user. The solution suggested also amalgamate the prospect of the voice control, via web interface or mobile application control. This enables it to adapt automatically with respect to the habits of the user and patterns of behaviour from tests evaluation the obtained results were produced, effectiveness and validation of the developed system. The study intended to bridge the gap of devices which have been developed which fail to function to the changing behaviour of users because when these devices are developed considerations of regularly changing habits of inhabitants were not considered. The study contributes to this systematic literature review by emphasizing the techniques that need to be considered to facilitate the automation of audio data inference in voice-activated systems.

An investigation of deep learning, as described in [16], was undertaken to advance the classification of encrypted traffic in smart speaker systems without the need for manual feature selection. The researchers in this study fed simple input features into the model, including interpacket time and information direction size, without the need for artificial feature extraction. The model achieved an accuracy surpassing 93%. The model surpasses various state-of-the-art models of voice command fingerprinting in a closed-world scenario. The researchers in this study conducted experiments in open-world scenarios, demonstrating the model's capability to effectively distinguish voice commands of interest to an attacker. They also illustrated the model's potential application in other domains, such as website fingerprinting.

The overall aim of this paper [17] was to investigate bandwidth efficient audio event detection technique and also noise robust. It is in the neural network scheme where the suggested denoising auto encoder – based completely joint scheme of deep neural network is realized. Extraction of audio features is conducted utilizing a Raspberry Pi edge device of computing which those to the local information server are fed for detection of audio events. At F1- score achieved of 90.6% of signal noise ratio 10db and accuracy of 90% a reflection of advanced categorization likelihood of noisy events of audio.

Arcades are described in this study [18] for extraction of context of representation of graphic of the system of home automation deep reinforcement learning is utilized and habitually updates its behaviour to that of the user. It is greatly understood when it comes to Convolutional Neural Network (CNN) that its requires large amount of data this prompted the researchers of this work to utilize CNN utilizing data generated artificially for feeding CNN with input which are raw for learning weights of CNN, execution of adapting decision model via reinforcement learning was performed, heterogenous information was presented and an approach of raw image was executed the decision model was based on deep learning. The results obtained by the study are as follows, Users trivial interchange (rewards) is utilized by the system for adaption of the process of decision making. For making a decision the graphical representation of heterogenous data multi model can be translated. The system

can adjust its behaviour to that of the user and sensors which are not installed, or which are faulty. Contextual data which is relevant can be learnt by the deep model and Online learning can be carried out utilizing restricted history, which is of the past, and absolutely no data pertaining to the future.

Table 2: Studies and inference methods identified from literature.

Author	**Inference Methods**
Li et al. [13]	Deep Learning
Berns et al. [14]	Gaussian Process
Filipe et al. [15]	Machine Learning
Mao et al. [16]	Deep Learning
Mondal et al. [17]	Deep Learning
Brenon et al. [18]	Deep Reinforcement Learning

The literature explored in this study reveals that the techniques used not only address automated inference but also offer an additional advantage of low latency in performance at a reduced cost. Deep Learning, Deep Reinforcement Learning, Adaptive Controller and Gaussian Process Classification could be utilized to translate heterogenous data and learn relevant contextual data in voice controlled smart homes. Additionally, literature review showed a solution [15] utilized the Online Learning techniques for a voice-activated smart home controller in order to obtain an intelligent solution capable of adapting and evolving with the user.

Conclusion and Future research

Standalone literature was utilized to gain insight into the existing literature regarding audio data inference in voice-activated systems. The entire literature review process, including literature search, extraction and analysis, was carried out to answer the research question. This study adopted a systematic review approach, which is appropriate for thorough reviews and provides proper methodology guidance. From the literature included in this study, various techniques were used to explore the literature, each yielding distinct results and contributing in different ways, as indicated in the table 1.

Deep neural networks were explored in the study, utilizing algorithms such as Bayesian Optimization and Deep Reinforcement Learning. The literature review analysis showed that these techniques can significantly outperform irrelevant baseline techniques that relied on randomness, simple statistics, or machine learning for creating dataset predictions related to the inference of speed and cost efficiency. Therefore, it is evident that deep neural networks and reinforcement learning can be considered for automating the inference of audio data in newly learned voice-activated systems.

References

[1] Sovacool, B.K. and Del Rio, D.D.F., 2020. Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies. *Renewable and sustainable energy reviews*, *120*, p.109663.
[2] Barker, S. and Parsons, D., 2022. Smart Homes or Real Homes: Building a Smarter Grid With "Dumb" Houses. *IEEE Pervasive Computing*, *21*(2), pp.100-104.
[3] Turpault, N., Wisdom, S., Erdogan, H., Hershey, J., Serizel, R., Fonseca, E., Seetharaman, P. and Salamon, J., 2020. Improving sound event detection in domestic environments using sound separation. *arXiv preprint arXiv:2007.03932*.
[4] Guirguis, K., Schorn, C., Guntoro, A., Abdulatif, S. and Yang, B., 2021, January. SELD-TCN: Sound event localization & detection via temporal convolutional networks. In *2020 28th European Signal Processing Conference (EUSIPCO)* (pp. 16-20). IEEE.
[5] Turpault, N., Serizel, R., Shah, A.P. and Salamon, J., 2019, October. Sound event detection in domestic environments with weakly labeled data and soundscape synthesis. In *Workshop on Detection and Classification of Acoustic Scenes and Events*.
[6] Tao, R., Yan, L., Ouchi, K. and Wang, X., 2021. Couple Learning for semi-supervised sound event detection. *arXiv e-prints*, pp.arXiv-2110.
[7] Martinek, R., Vanus, J., Nedoma, J., Fridrich, M., Frnda, J. and Kawala-Sterniuk, A., 2020. Voice communication in noisy environments in a smart house using hybrid LMS+ ICA algorithm. *Sensors*, *20*(21), p.6022.
[8] Shafi, A., Saeed, S., Bamarouf, Y.A., Iqbal, S.Z., Min-Allah, N. and Alqahtani, M.A., 2019. Student outcomes assessment methodology for ABET accreditation: A case study of computer science and computer information systems programs. *IEEE Access*, *7*, pp.13653-13667.
[9] López-Pernas, S., Saqr, M. and Apiola, M., 2023. Scientometrics: a concise introduction and a detailed methodology for mapping the scientific field of computing education research. *Past, Present and Future of Computing Education Research: A Global Perspective*, pp.79-99.
[10] Ferreira, D.J., Mateus-Coelho, N. and Mamede, H.S., 2023. Methodology for Predictive Cyber Security Risk Assessment (PCSRA). *Procedia Computer Science*, *219*, pp.1555-1563.
[11] Mamatova, Z.H., 2023. CONCEPTUAL AND METHODOLOGICAL FOUNDATIONS OF IMPROVING COMPUTER SCIENCE TRAINING IN THE CONDITIONS OF DIGITAL TECHNOLOGY. *European International Journal of Pedagogics*, *3*(01), pp.42-54.
[12] Shafiezadeh, S., Duma, G.M., Mento, G., Danieli, A., Antoniazzi, L., Del Popolo Cristaldi, F., Bonanni, P. and Testolin, A., 2023. Methodological Issues in Evaluating Machine Learning Models for EEG Seizure Prediction: Good Cross-Validation Accuracy Does Not Guarantee Generalization to New Patients. *Applied Sciences*, *13*(7), p.4262.
[13] Li, Y., Han, Z., Zhang, Q., Li, Z. and Tan, H., 2020, July. Automating cloud deployment for deep learning inference of real-time online services. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (pp. 1668-1677). IEEE.
[14] Berns, F., Hüwel, J. and Beecks, C., 2022. Automated Model Inference for Gaussian Processes: An Overview of State-of-the-Art Methods and Algorithms. *SN Computer Science*, *3*(4), p.300.
[15] Filipe, L., Peres, R.S. and Tavares, R.M., 2021. Voice-activated smart home controller using machine learning. *IEEE Access*, *9*, pp.66852-66863.
[16] Mao, J., Wang, C., Guo, Y., Xu, G., Cao, S., Zhang, X. and Bi, Z., 2022. A novel model for voice command fingerprinting using deep learning. *Journal of Information Security and Applications*, *65*, p.103085.
[17] Mondal, S. and Barman, A.D., 2022. Deep learning technique based real-time audio event detection experiment in a distributed system architecture. *Computers and Electrical Engineering*, *102*, p.108252.
[18] Brenon, A., Portet, F. and Vacher, M., 2018. Arcades: A deep model for adaptive decision making in voice controlled smart-home. *Pervasive and Mobile Computing*, *49*, pp.92-110.

Authors Biography

Vuyolwethu Sunday Mantiyane, is currently in quest of Masters of Computer Science degree at University of Fort Hare. He obtained his undergraduate degree which is, Bachelor of Science in Computer Science and Geographic Information System (GIS) in 2021 from University of Fort Hare and was further awarded a Bachelor of Science Honours degree in Computer Science in 2022 from University of Fort Hare. His contemporary interests in research comprise of, Machine Learning and Sensor Networks.

Phumzile Nomnga, is a Lecturer in the Department of Computer Science at the University of Fort Hare and is concurrently pursuing a PhD in the same institution. His research focus encompasses several areas, notably the Internet of Things (IoT), Artificial Intelligence (AI), Mobile Networks, and Big Data Analytics.