Using Finite State Machines with Gesture Recognition Model for Context-Aware Decision in Automotive Applications

A Thesis Project

Presented to the Faculty of the

Department of Computer, Information Sciences and Mathematics

University of San Carlos

In Partial Fulfillment

of the Requirements for the

CIS 3202N – Automata Theory and Formal Languages

Members:

Rosales, Jade Andrie

Samson, Jose Glen

Stewart, Christian Anthony

May 18, 2023

ABSTRACT

Gesture recognition technology is a new and emerging technology in the automotive industry that has the potential to enhance the driving experience and safety of drivers. To address challenges such as low accuracy and reliability, limited recognition range, and inability to integrate seamlessly into the existing automotive system, a context-aware gesture recognition system using Long Short-Term Memory (LSTM) with Finite State Machine (FSM) models is proposed. The proposed system uses a dataset that is created by the researchers, which is pre-processed to remove any extraneous noise or interference, and to segment it into meaningful gestures. The pre-processed data is then used to train and test the FSM gesture recognition model in automotive applications. The model's performance is then evaluated based on metrics such as accuracy, precision, and recall. The proposed system aims to recognize a wide range of gestures accurately, making it more personalized and user-friendly, leading to improved safety, convenience, and comfort for drivers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Rationale of the Study

In the automotive industry, there is an increasing trend of integrating advanced technology to enhance the driving experience and safety of drivers according to Kamran et al. (2022). Gesture recognition technology is an emerging technology that can make driving safer, easier, and more convenient by allowing drivers to perform various functions, such as adjusting the volume or answering calls, without taking their eyes off the road or hands off the wheel (Fong et al. 2001). However, the existing gesture recognition systems face several challenges, such as low accuracy and reliability, limited recognition range, and inability to integrate seamlessly into the existing automotive system.

The proposed context-aware gesture recognition system approach using Finite State Machine (FSM) models with Long Short-Term Memory (LSTM) can overcome these challenges by accurately recognizing a wide range of gestures and integrating seamlessly into the existing automotive system. By integrating FSM with LSTM, a context-aware model can be developed that can recognize the context in which the gesture is performed to improve the accuracy and reliability of the gesture recognition system.

The proposed gesture recognition system can be used for various functions, such as adjusting the volume, changing the music track, answering or rejecting calls, controlling the air conditioning system, and activating the navigation system. These functions can be performed by the driver with simple gestures, such as a play, next, or previous, without taking their eyes off the road or hands off the wheel. Moreover, the proposed system can be customized to the driver's preferences and driving style, making it more personalized and user-friendly.

The proposed system can have several benefits for drivers, such as improved safety, convenience, and comfort. By enabling drivers to perform various functions without taking their hands off the wheel or eyes off the road, the system can reduce distractions and improve situational awareness, leading to a safer driving experience. The proposed system can also reduce the cognitive load on the driver, making driving more comfortable and less stressful.

1.2 Statement of the Problem

1.2.1 General Objective

This study aims to develop a context-aware gesture recognition application for automobiles using a Finite State Machine (FSM) with Long Short-Term Memory (LSTM) Model.

1.2.2 Specific Objective

1. Create a dataset to cater specifically for automotive application gesture recognition.
2. Develop an automata-based pattern recognition algorithm for gesture recognition applications.
3. To develop a FSM with LSTM model for gesture recognition.
4. Evaluate the model using standard performance metrics.

1.3 Significance of the Study

The outcome of this study may be instrumental to the following parties:

General Vehicle Drivers. Firstly, it would benefit drivers who would be able to use hand gestures to control various functions in their cars, such as changing the music, adjusting the temperature, or answering a call without having to take their hands off the steering wheel. This would improve safety on the road by reducing distractions and increasing focus on driving.

Automobile Manufacturers. Firstly, it can enhance the overall driving experience by providing a more intuitive and seamless interaction between the driver and the vehicle. This can result in increased customer satisfaction and loyalty. Secondly, it can improve safety by allowing drivers to control various functions without having to take their hands off the steering wheel or eyes off the road. This can reduce the likelihood of accidents caused by distractions or delayed reaction times. Finally, the technology can give automobile manufacturers a competitive edge in the market by offering unique and innovative features that differentiate their products from competitors.

Researchers. The study is a novel and innovative research direction. It addresses the limitations of existing methods, particularly in distinguishing between different gestures and interpreting their meanings. The research could lead to a better understanding of the underlying mathematical principles of gesture recognition and inspire further exploration and development of related algorithms.

Future Researchers. The study could serve as a foundation for future research in the field. It provides an alternative to existing methods for pattern recognition and opens up possibilities for further exploration and development of automata-based algorithms. Future researchers could build upon the findings and insights to advance the field of gesture recognition and its applications.

1.4 Scope and Limitations

The scope of this thesis research study on using FSM for gesture recognition for automotive applications coupled with LSTM is focused on the development and implementation of a reliable and accurate system for gesture recognition in automobiles. The study will explore the potential of using a finite state machine and a deep learning model with a long short-term memory model to detect and interpret hand gestures performed by drivers in a moving vehicle. The system aims to provide a safer and more convenient way for drivers to control various in-car functions, such as music playback, navigation, and air conditioning, without the need for physical buttons or touch screens.

However, there are also limitations to consider in this research study. Gestures are handpicked or specifically chosen by researchers applicable for automotive application usage. Dataset is to be made from scratch which will involve manpower and planning. The researchers will take careful steps to ensure that the dataset is representative of a variety of potential scenarios in which the gesture recognition system may be used, while also acknowledging the limitations of the dataset they have created.

CHAPTER 2
REVIEW OF RELATED LITERATURE

This chapter presents a list of research done with due diligence by others that provide significant insight into the researchers' study. This includes methods that have been researched and applied which have provided the foreground for the researchers' thought processes and methodologies. The publications stated below may or may not be directly related to the current research, but they provide crucial information that supports the study and provide the general gist or background upon which the study is built.

Finite State Machine

Finite State Machines (FSMs) are mathematical models used to represent and analyze systems with discrete and sequential behaviors. They consist of a finite set of states, a set of inputs, a set of outputs, and a set of transitions that describe how the system transitions between states based on inputs and produces outputs. FSMs are widely used in various domains, including computer science, control systems, robotics, software engineering, and artificial intelligence.

At any given time, an FSM is in a specific state, and it receives inputs from its environment. These inputs trigger transitions that cause the FSM to move from one state to another. The transitions are defined by conditions associated with the inputs, and they specify the actions or outputs to be taken when a transition occurs. FSMs can operate in different modes, such as deterministic (where each input uniquely determines the next state) or non-deterministic (where multiple transitions can be enabled simultaneously).

FSMs are particularly useful for modeling systems with discrete behaviors and a finite number of possible states. They provide a structured approach to understanding and designing complex systems by breaking them down into a series of well-defined states and transitions. FSMs enable the analysis of system behavior, the detection of errors or inconsistencies, and the synthesis of correct and efficient control strategies.

The book "Elements of Robotics" by Ben-Ari and Mondada (2017) provides a comprehensive exploration of the fundamental concept of Finite State Machines (FSMs) and their applications in various domains. The book begins by introducing the basic concepts of FSMs, emphasizing their role in modeling systems with discrete behaviors. It then goes beyond the theoretical aspects of FSMs and delves into their practical applications across a wide range of disciplines. The authors also discuss advanced topics related to FSMs, such as non-deterministic FSMs, finite state transducers, and extended FSM models. The book's clarity and comprehensiveness makes it an indispensable reference for researchers, engineers, and students in the fields of computer science, robotics, and systems engineering. Through this comprehensive exploration, readers gain a deeper appreciation for the versatility and power of FSMs as a modeling and analysis tool, enabling them to advance their own research and contribute to the development of innovative solutions in various domains (Ben-Ari & Mondada, 2017).

Using Finite State Machines (FSM) for Gesture Recognition

In the paper of Hong, Turk, and Huang (2000), it presents an efficient approach for modeling and recognizing hand and mouse gestures using Finite State Machines (FSM). The proposed method allows for a semi-automatic way of constructing gesture models and does not require large datasets for training. The algorithm first decouples spatial and temporal information of the data, learns spatial information for data segmentation and alignment, and then learns temporal information from the aligned data segments. The FSM recognition procedure is incorporated with the Knuth-Morris-Pratt (KMP) algorithm, a fast string-matching algorithm, to achieve fast recognition speed. The approach was tested on hand and mouse gestures, achieving recognition rates of 90% or better for hand gestures and 70-100% for mouse gestures. The results demonstrate the potential for this approach to be applied to a large vocabulary. Overall, the paper presents a promising approach for gesture recognition with potential applications in various fields.

Johnston and Bangalore (2000) conducted a research paper on the application of Finite-state Machines (FSMs) in the context of multimodal parsing and understanding. The paper highlights the importance of multimodal parsing in capturing the intricacies of human communication, and proposes a framework that allows for the efficient representation and processing of multimodal input. The authors employ a combination of Finite-state Transducers (FSTs) and Finite-state Automata (FSAs) to model and process various modalities, including speech and gesture, in a synchronized manner (Johnston & Bangalore, 2000). Through their experimental evaluations, the authors demonstrate the effectiveness of their approach in achieving accurate multimodal parsing and understanding. The research concludes that Finite-state Multimodal Parsing and Understanding provides a promising framework for capturing and interpreting multimodal input in human-computer interaction scenarios. The findings of this research provide a solid foundation for further exploration and development of multimodal parsing techniques, opening up new avenues for natural and intuitive human-computer interaction.

Aksaç et al. (2011) explored the application of a Finite State Machine (FSM) combined with distance classifiers for real-time hand posture and gesture recognition, specifically in the context of virtual mouse operations. They proposed a novel approach that combines distance classifiers and an FSM-based framework to recognize and interpret hand postures and gestures in real-time, allowing users to control a virtual mouse with hand movements. The paper presents the implementation details of their proposed system, which involves the use of depth data from a depth sensor to capture hand movements. Through extensive experiments and evaluations, Aksaç, ztürk, and zyer demonstrate the effectiveness of their approach in achieving real-time and accurate hand posture and gesture recognition. The research concludes that the combination of distance classifiers and FSM-based recognition provides a robust and efficient framework for real-time hand posture and gesture recognition. The findings of this study have practical implications for the development of intuitive and natural user interfaces, furthering the progress of virtual mouse operations and related applications.

Using Fuzzy Logic for Gesture Recognition

A fuzzy rule-based method for the recognition of hand gestures was introduced in the paper of Bedregal, Costa, and Dimuro (2006). The method uses the set of angles of finger joints for the classification of hand configurations, and classifications of segments of hand gestures for recognizing gestures. There are 27 possible finger configurations that are considered. Each of these configurations are codified into variables to indicate that the three joints are either straight or curved. The hand configuration can also be determined by considering each finger configuration. Then finally using fuzzy logic to determine the segmentation of the gesture in monotonic hand segments. The set of all lists of segments of a given set of gestures determine a set of finite automata, which are able to recognize every such gesture. The paper shows promising results and is suitable for the application to the recognition of hand gestures of sign languages.

Another paper that follows this method is the paper of Verma and Dev (2009). The paper proposes an approach for hand gesture learning and recognition that combines finite state machines and fuzzy logic. The researchers use edge detection and vector extraction to identify the location of the user's hands in 2D image positions. The information of spatial and temporal domain is first separated. The data is grouped into clusters based on temporal alignment using Fuzzy c-mean clustering which allows one piece of data to belong to two or more clusters. The clusters of hand postures are used to determine the states of the finite state machine(s), which are used to match the succeeding gesture. The number of states/clusters represents a trade-off between the accuracy of recognizing the gesture and the amount of special/temporal data. The technique is demonstrated to be successful for a set of gestures such as waving left/right hand, signaling to stop, forward, rewind, etc. The approach has potential for further development and applications in various fields.

Using Finite State Machine with Hidden Markov Model (HMM) for Gesture Recognition

A study by P. Hong, M. Turk, and T. S. Huang (2000) proposes a novel method of gesture recognition using finite state machines. The study focuses on recognizing hand gestures in real-time by modeling them as FSMs. The researchers propose a hierarchical FSM architecture that can recognize complex gestures by combining basic gesture FSMs. The study also introduces a novel method for modeling gesture dynamics using hidden Markov models (HMMs). To evaluate the proposed method, the researchers conducted experiments on two gesture datasets, one containing basic hand gestures and the other containing more complex gestures. The results showed that the proposed method achieved high recognition accuracy for both datasets. Overall, the study presents a promising approach to gesture recognition using FSMs and HMMs. The proposed method has potential applications in various fields, including human-computer interaction, virtual reality, and robotics.

The paper of Elmezain et al. (2008) proposes an automatic system for real-time recognition of isolated and continuous hand gestures for Arabic numbers 0-9 using HMM. The system applies HMM using different topologies and numbers of states ranging from 3 to 10 for isolated gestures, and zero-codeword detection with static velocity motion for continuous gestures. The proposed system uses orientation dynamic features obtained from spatio-temporal trajectories and codewords. The system achieves an average recognition rate of 98.94% and 95.7% for isolated and continuous gestures, respectively, with the LeftRight Banded topology with 5 states presenting the best performance. The researchers suggest future work on motion trajectory carried out by fingertip instead of hand centroid point using a multi-camera system over combined features. Overall, the proposed system shows promising results for real-time recognition of hand gestures for Arabic numbers.

Using Convolutional Neural Network (CNN) for Gesture Recognition

In the paper of Lin, Hsu, and Chen (2014), it proposes a convolutional neural network (CNN) approach to hand gesture recognition. The paper describes a system that is capable of detecting and recognizing hand gestures in real-time using a camera-based input. The system consists of three main components: image acquisition, feature extraction, and classification. The image acquisition component captures hand images using a webcam, and the feature extraction component extracts features from the images using CNNs. They also adopted a Gaussian Mixture model (GMM) to robustly filter out non-skin colors of an image. The model was trained to learn seven gesture types and got results that showed that the average recognition rates were around 99%. However the current system still needs enhancement to the recognition capability for complex human tasks.

A method based on multimodal data fusion and multiscale parallel CNN is proposed in the paper of Gao, Liu, and Ju (2021). The data fusion is conducted on the sEMG signal, the RGB image, and the depth image of hand gestures which are then fused to generate two different scale images by downsampling. The data is then inputted into two subnetworks of the parallel CNN to obtain two hand gesture recognition results. The results are then combined to obtain the final hand gesture recognition result. Ten commonly used HRI (Human-Robot Interaction) hand gestures are designed. As a result, the system's accuracy for the 10 HRI hand gestures is greater than or equal to 78%. The system's limitations include its inability to recognize dynamic hand gestures. Dynamic hand gestures are more practical and difficult than static hand motions hence why the system is less accurate in this field.

The paper of Islam, M. Z. et al. (2019), proposes a static hand gesture recognition method using CNN. The researcher applied data augmentation techniques to the dataset and trained the model on 8000 images, achieving an accuracy of 97.12% on the test set of 1600 images across 10 classes. The results show that data augmentation has a significant impact on the accuracy of the model. The researchers suggest some potential areas of future work, including the use of knowledge-driven methodologies to improve accuracy and the recognition of gestures made with both hands. These suggestions are useful and provide a direction for further research in this area. Overall, the paper demonstrates the effectiveness of CNNs and data augmentation techniques.

Using FSM with 3DCNN and LSTM for Gesture Recognition

Hakim et al. (2019) proposed a dynamic hand gesture recognition model that combines 3D convolutional neural network (CNN) and long short-term memory (LSTM) with a finite state machine (FSM) context-aware model. The model aims to improve the accuracy and robustness of hand gesture recognition by incorporating temporal information and context-awareness. The 3D CNN is used to extract spatial and temporal features from the input gesture sequences, while the LSTM is used to model the temporal dynamics of the gestures. The FSM context-aware model is used to model the transitions between gestures and incorporate contextual information such as the user's hand position and orientation.

To extract the spatio-temporal information, a three-dimensional convolutional neural network (3DCNN) and a long short-term memory (LSTM) model were combined. Finite State Machine (FSM) communicates the model to regulate the class determination outcomes based on application context at the conclusion of the classification. The results showed that, for eight selected motions, the combination of depth and RGB data had a 97.8% accuracy rate, while the FSM increased the identification rate in real-time from 89% to 91%.

The authors first collected a dataset of dynamic hand gestures performed by multiple users in various lighting conditions and viewpoints, using a Kinect sensor. They then preprocessed the data by converting it to 3D voxel grids and normalizing the hand size and position. The preprocessed data was then fed into a 3DCNN to extract spatiotemporal features from the hand movements. The output of the 3DCNN was then fed into an LSTM network to capture the temporal dependencies and memory-based modeling of the gestures. Finally, the output of the LSTM was fed into an FSM context-aware model, which used a set of rules to constrain the possible transitions between gestures and ensure the consistency and coherence of the recognized gestures.

The authors evaluated the performance of their approach using several metrics, including accuracy, precision, recall, and F1 score. They compared their approach with several state-of-the-art methods, such as HMMs, SVMs, and CNNs, and showed that their approach achieved superior results in terms of accuracy and robustness. They also conducted several experiments to analyze the impact of various factors, such as the number of hidden units in the LSTM, the size of the FSM state space, and the effects of noise and occlusion in the input data. The results showed that their approach was able to handle various challenges and achieve high performance under different conditions.

The approach proposed by Hakim et al. (2019) involves the use of several components that can be viewed as automata or models. The 3D convolutional neural network (3DCNN) can be viewed as a type of automata that recognizes patterns in the spatial and temporal domain. The long short-term memory (LSTM) network can be seen as a type of automata that captures the temporal dependencies and memory-based modeling of the gestures. The finite state machine (FSM) context-aware model can be seen as a type of automata that enforces constraints on the possible transitions between gestures and ensures the consistency and coherence of the recognized gestures. Therefore, the proposed approach involves the integration of multiple automata or models to achieve higher accuracy and robustness.

Moreover, the study by Hakim et al. (2019) highlights some of the challenges and limitations of existing methods, such as the reliance on 2D images or depth maps and the inability to capture the temporal dynamics and contextual information of the gestures. These challenges are also relevant to automata theory, as the ability to recognize or generate languages depends on the ability to capture the spatial and temporal patterns and the contextual constraints of the language. Therefore, the proposed approach can be seen as a novel way to overcome some of these challenges and limitations and advance the field of automata theory.

Utilizing TensorFlow Framework for Gesture Recognition

The research study by Zeng et al. (2018) on the design of a CNN model for gesture recognition based on the TensorFlow framework is an important contribution to the field of computer vision and machine learning. The study presents a novel approach to recognizing hand gestures that can be used in various real-world applications, such as human-computer interaction, robotics, and virtual reality. The authors propose a CNN-based model that achieves high accuracy and robustness in recognizing a variety of hand gestures.

The study by Zeng et al. (2018) demonstrates the effectiveness of using a CNN-based model for gesture recognition compared to other traditional machine learning methods, such as decision trees and SVMs. The authors conducted experiments on a large dataset of hand gesture images and showed that their CNN-based model outperformed the other methods in terms of accuracy and speed. Moreover, the study provides a detailed analysis of the CNN model's architecture and parameters and highlights the importance of optimizing these parameters for achieving the best performance.

The findings of the study by Zeng et al. (2018) have important implications for the development of real-world applications that involve gesture recognition. The proposed CNN-based model can be used in various scenarios, such as sign language recognition, game control, and robot navigation. The high accuracy and robustness of the model can improve the overall user experience and enable more intuitive and natural interaction between humans and machines. The authors also recommend further research in exploring the use of deep learning techniques, such as recurrent neural networks and attention mechanisms, to enhance the performance of the model and enable more complex and subtle gesture recognition.

Overall, the research study by Zeng et al. (2018) presents a novel approach to gesture recognition based on a CNN model and provides valuable insights into the design and optimization of such models. The study demonstrates the effectiveness of using CNNs for recognizing hand gestures and highlights the importance of optimizing the model's architecture and parameters. The findings and recommendations of the study can inform the development of real-world applications that involve gesture recognition and stimulate further research in the field of computer vision and machine learning.

CHAPTER 3
DESIGN AND METHODOLOGY

3.1 Research Instrument or Sources of Data

The dataset used for training and testing the FSM Context-Aware with LSTM model for gesture recognition in automotive applications will be created by the researchers themselves. The researchers will be responsible for designing and performing the specific gestures to be recognized by the system. By creating their own dataset, the researchers can ensure that the gestures are accurately and consistently performed, reducing the potential for variability in the data.

3.2 Research Procedure

3.2.1 Gathering of Data

1. Identify the specific gestures to be used in the automotives through extensive research.
2. Set up a recording environment that allows for consistent and reliable capture of the gestures.
3. Gesture performing will be done by researchers themselves.
4. Instruct how to perform the gestures and ensure that they are able to do so consistently and accurately. This may involve providing visual examples or demonstrations, as well as providing feedback and coaching to improve performance.
5. Record the gestures performed by researchers using the chosen recording environment. Ensure that the data is properly labeled and organized to facilitate analysis and training of the gesture recognition model.

3.2.2 Treatment of Data

The collected data will be pre-processed to remove any extraneous noise or interference, and convert it into a format that can be used by the gesture recognition algorithm. This may involve standardizing the format of the data, such as converting it into a specific file type or compressing it for storage.

The data will the classified into the following gestures:

Table 1

Gesture Classifications

id	**Label**	**Gesture**
1	*D0X*	*Non-gesture*
2	G01	Volume Up
3	G02	Volume Down
4	G03	Mute
5	G04	Unmute
6	G05	Play
7	G06	Pause
8	G07	Back
9	G08	End
10	G09	Next
11	G10	Previous
12	G11	Home

The classification of gestures is a critical step that leads to the successful recognition of these gestures in subsequent stages of the study.

3.3 Conceptual Framework

The goal of this study involves the integration of FSM with a GRM to develop a context-aware model for automotive applications aims to capture the interaction between the user, the camera capturing the gestures, the GRM for recognizing the gestures, the FSM for managing the context and decision-making, and ultimately generating relevant results.

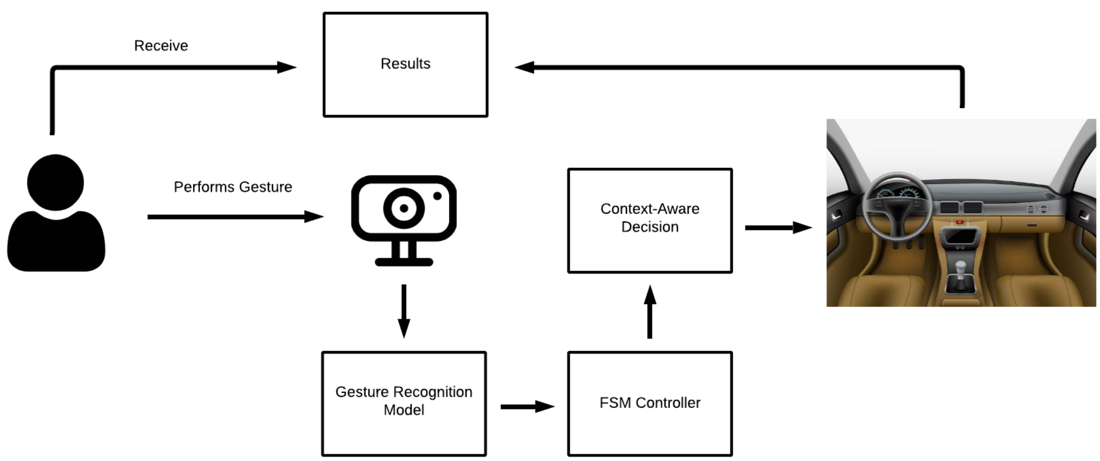

Figure 1. Conceptual Framework of the Study

The following figure above shows the interaction between the user and the proposed system. Similarly, it provides a general overview of the system or model on how it should be utilized as the researchers intended.

The user in this case, engages with the automotive vehicle by performing gestures, which are captured by a camera. The captured visual data is then analyzed by the GRM, employing computer vision and machine learning techniques to recognize the gestures accurately. The recognized gestures are passed on to the FSM, which acts as the core component of the context-aware model.

The FSM manages the context based on predefined states and transitions, taking into account the current state, previous states, and the recognized gestures as inputs. Using this contextual information, the FSM makes context-aware decisions, considering factors such as driving conditions, vehicle status, and user preferences. These decisions lead to specific results or actions, triggering relevant functions or controls within the automotive vehicle based on the recognized gestures and the current context.

3.4 Analysis and Design

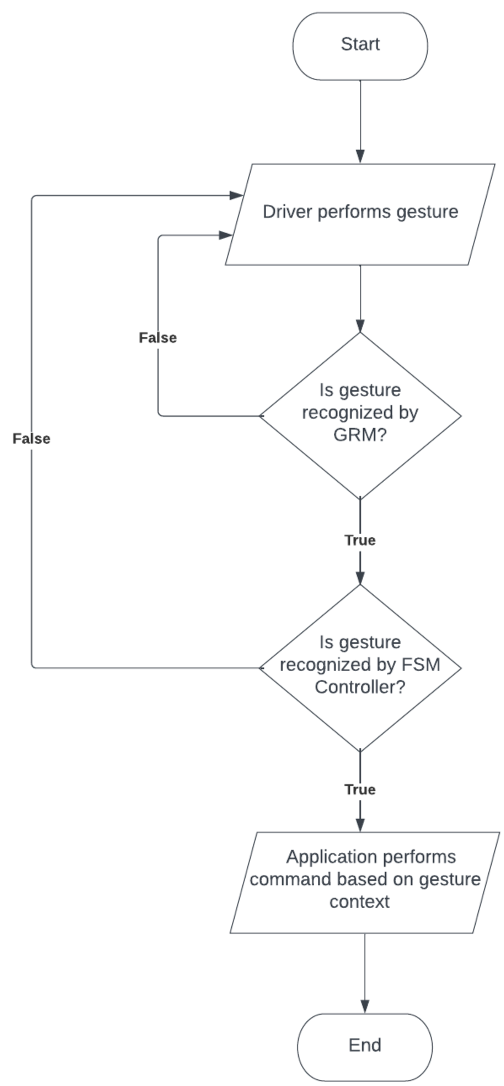

Figure 2. Flowchart of the context-aware model

Figure 2 provides an overview of the context-aware model's flowchart. It outlines the sequential steps involved in the decision-making process. It starts by acquiring the input data captured by a camera. The input data in question is the gesture performed by the driver.Next, it will be fed to the GRM for recognition. Based on the recognized gesture and the determined context, the model transitions to an appropriate state in the Finite State Machine (FSM). The FSM represents different states that the system can be in, each corresponding to a specific set of actions or decisions.Within each state, the context-aware model evaluates the available options and makes a decision considering both the recognized gesture and the current context.After the context-aware decision is made, the model generates appropriate outputs, such as control signals to adjust vehicle settings or visual/audio cues to inform the driver. These outputs are designed to enhance the overall driving experience and safety.

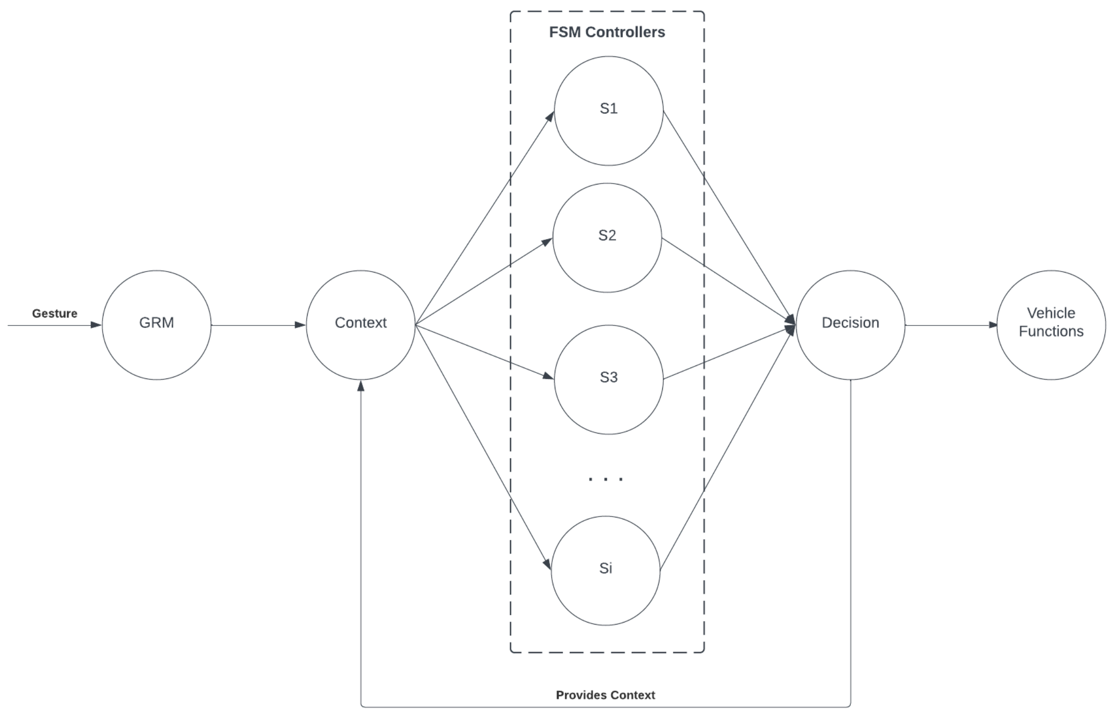

Figure 3. GRM with FSM controller models

Figure 3 provides a detailed illustration of the GRM with FSM controller models. This figure depicts the integration of the GRM and FSM, showcasing their interconnection and interaction. The GRM and FSM controllers are tightly integrated, enabling seamless communication and data exchange. The recognized gesture from the GRM serves as input to the FSM controller, influencing the state transitions and subsequent decision-making.

The figures below will show the Finite State Machines (FSM) controller model for the Main Menu, Temperature Control, and Voice Call Prompt commands. It shows the state transition diagram of the proposed gesture recognition model for automotive applications. A mealy machine will be employed for all models due to its flexibility in generating outputs based on input conditions. It can also respond immediately to input changes and produce output accordingly which is ideal for the study since the model needs to process the gestures in real time.

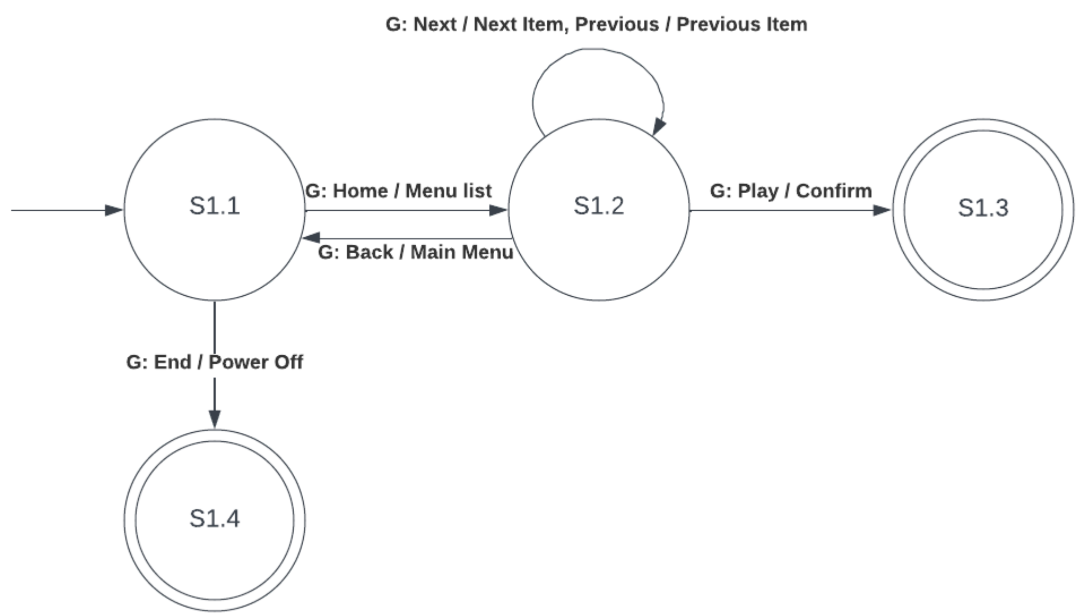

Figure 4. FSM Controller Model for Main Menu

Starting with Figure 4, the application will go to the menu list once a “Home” gesture. In the menu list, the driver will be allowed to go to the next or previous item by performing the “Next” and the “Previous” gesture which they can confirm by doing the “Play” gesture. The driver can also choose to power off the application by performing the “End” gesture.

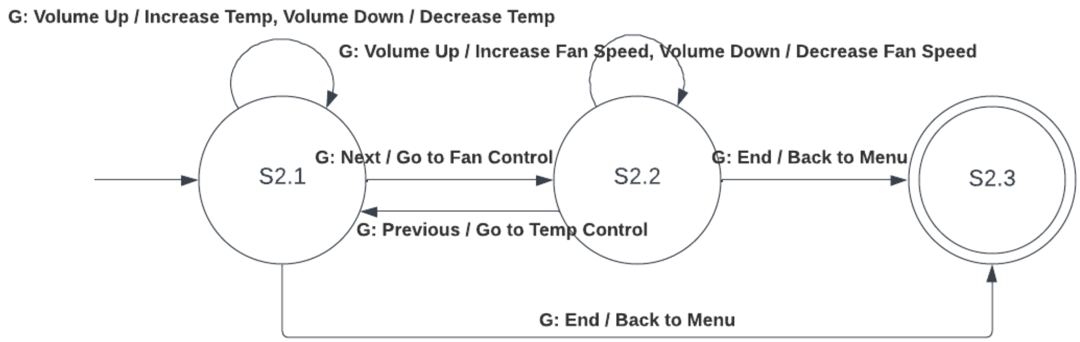

Figure 5. FSM Controller Model for Temperature Control

Next, in Figure 5, the driver will be able to increase/decrease the temperature using the “Volume Up/Down” gesture. By performing the “Next” and “Previous” gesture, the driver can toggle between the Temp Control and Fan Control settings respectively. The fan speed can also be controlled with the same gestures as the Temp Control. Like in the previous figure, the driver may choose to go back to the menu by performing the “End” gesture.

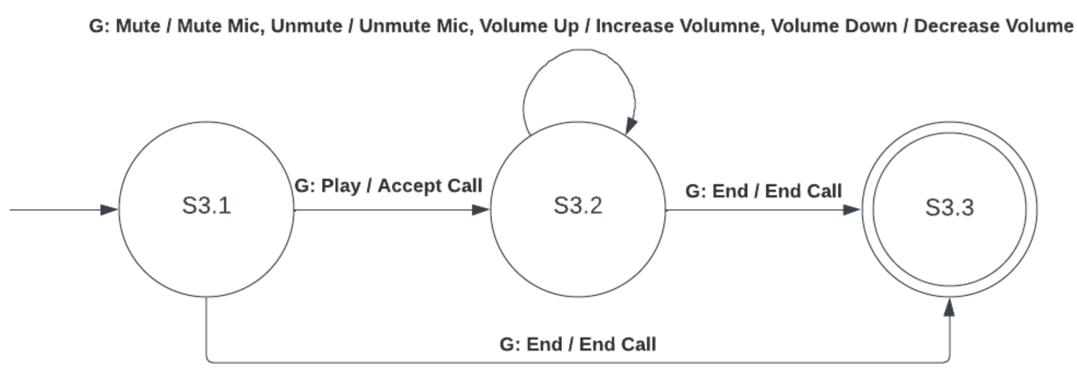

Figure 6. FSM controller model for Voice Call Prompt

Lastly, in Figure 6, the driver will be able to accept an incoming call by performing the “Play” gesture after which he is given the functions to mute mic, unmute mic, increase volume, and decrease volume by performing the gestures: “Mute”, “Unmute”, “Volume Up”, “Volume Down” respectively. And also the driver may choose to end the call by performing the “End” gesture.

3.5 Development Model

The Kanban Agile Framework is a powerful approach that promotes lean management and continuous improvement in any given work system. Its unique emphasis on visualizing work processes, setting work-in-progress limits, and maximizing efficiency makes it an ideal method for conducting research studies, especially in implementing a conceptual framework. With Kanban's adaptability to change and ability to prioritize tasks based on importance and urgency, it provides researchers with greater flexibility and control over their study's progress (Oren, 2023).

Incorporating Kanban into the research methodology allows for a more streamlined workflow, resulting in improved productivity and a more efficient use of time and resources (McDonald, 2023). Through the use of visual aids, such as boards and cards, researchers can easily track progress and identify any potential bottlenecks that may arise. This enables them to quickly take action and adjust their approach as needed, ensuring that the study is completed on time and within budget, while maintaining a high level of accuracy and quality. Overall, the Kanban Agile Framework is an essential tool for any researcher looking to optimize their study's efficiency and effectiveness (Oren, 2023).

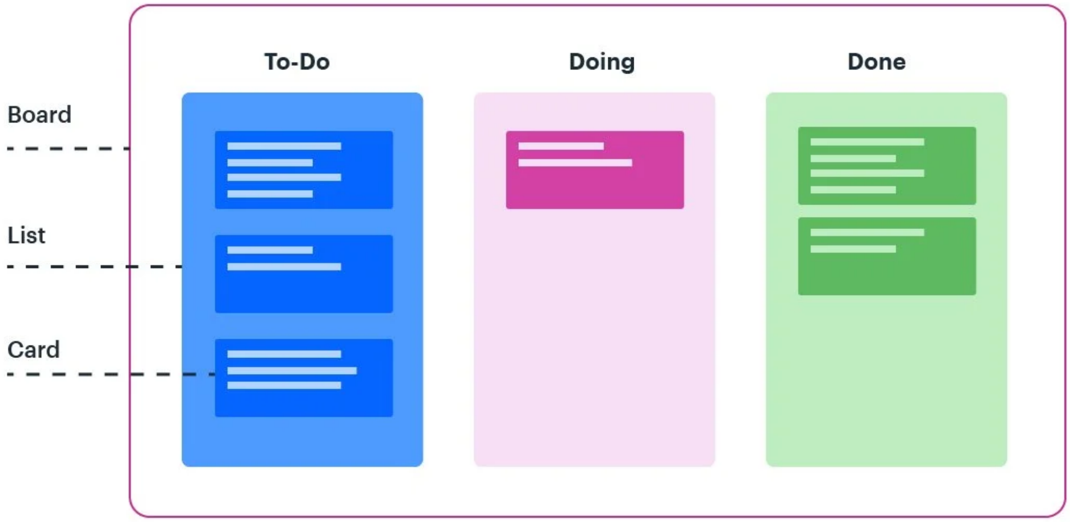

Figure 7. Kanban Agile Framework

Figure 5 shows the Kanban board consists of three main divisions: To-Do, Doing, and Done. All three columns are essential components of a Kanban board. These columns allow for a clear and concise visualization of the workflow and provide a framework for tracking progress.

The To-Do column serves as a repository for all the tasks that need to be completed, and these tasks are typically ordered based on their priority.

Once a task is selected to be worked on, it is moved to the Doing column, where it is actively being worked on. The Doing column usually has a limit on the number of tasks that can be in progress at any given time to prevent team members from becoming overloaded.

Finally, when a task is completed, it is moved to the Done column, indicating that it has been finished. This visual representation of the workflow helps to ensure that team members remain focused on completing tasks and can easily track their progress, leading to greater productivity and efficiency.

3.6 Development Approach

The bottom-up approach will be employed in the study due to its suitability for the nature of the problem being addressed. Specifically, the study aims to identify and recognize specific gestures that can be used to control different functions in automobiles. Given the complex and nuanced nature of hand gestures, it is likely that the most effective way to identify and recognize them is by analyzing and combining simpler building blocks of gesture data. For instance, the shape and movement of individual fingers or hand positions and orientations can be combined to create more complex gestures.

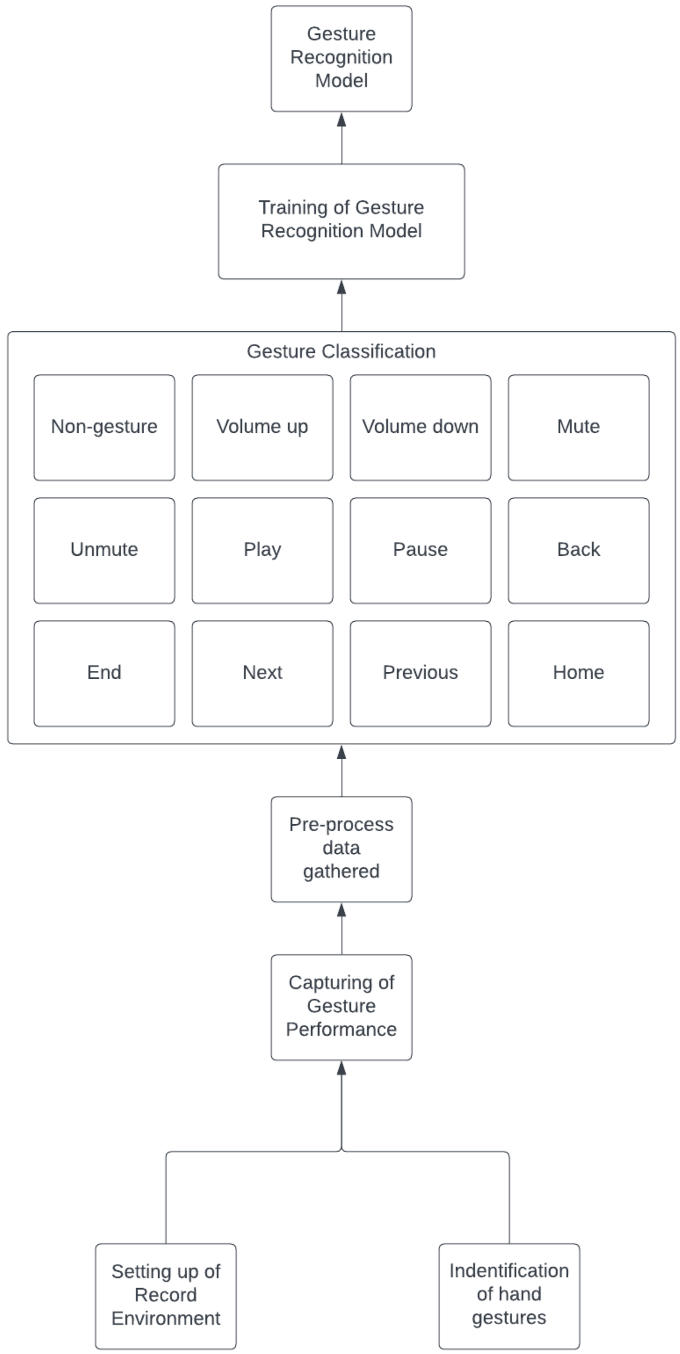

Figure 8. Bottom-Up Approach of Developing the Model

The bottom-up approach allows the researchers to start with these simple building blocks of gesture data and iteratively combine them to identify and create more complex gestures that are appropriate for use in automobiles.

Moreover, the bottom-up approach is well-suited for the iterative and exploratory nature of the research. The study is expected to involve extensive research and experimentation to identify and refine the appropriate gestures for use in automobiles. Starting with a few high-level concepts and pre-planning, as would be done in a top-down approach, may limit the ability of the researchers to explore and discover new and effective gesture combinations. By starting with the simpler building blocks of gesture data and iteratively combining them, the researchers can explore a wider range of potential gestures and refine their understanding of how these gestures can be used effectively in the context of automobiles.

3.7 Verification, Validation, and Testing

Once the primary goal of the study, which is to develop a gesture recognition application for automotive applications using LSTM with FSM Context-Aware Model, is accomplished, it is essential to verify and validate the system to ensure its integrity. This chapter's subsection focuses on the validation and verification procedures used for assessing the accuracy of the outcomes and metric scores, including precision, recall, and F1, as well as the procedures utilized for training and testing the dataset.

To validate and verify the system, performance measures such as K-fold cross-validation will be employed. K-fold cross-validation involves dividing the dataset into K subsets, each of which serves as a testing set and is compared to the other subsets to accurately assess the model's correctness. For this study, we will be using tenfold cross-validation, and the accuracy, precision, recall, and F1 score of each test will be averaged to determine the model's average metrics. The use of these performance measures will ensure that the developed gesture recognition system for automotive applications is robust, accurate, and effective in detecting and interpreting driver's gestures.

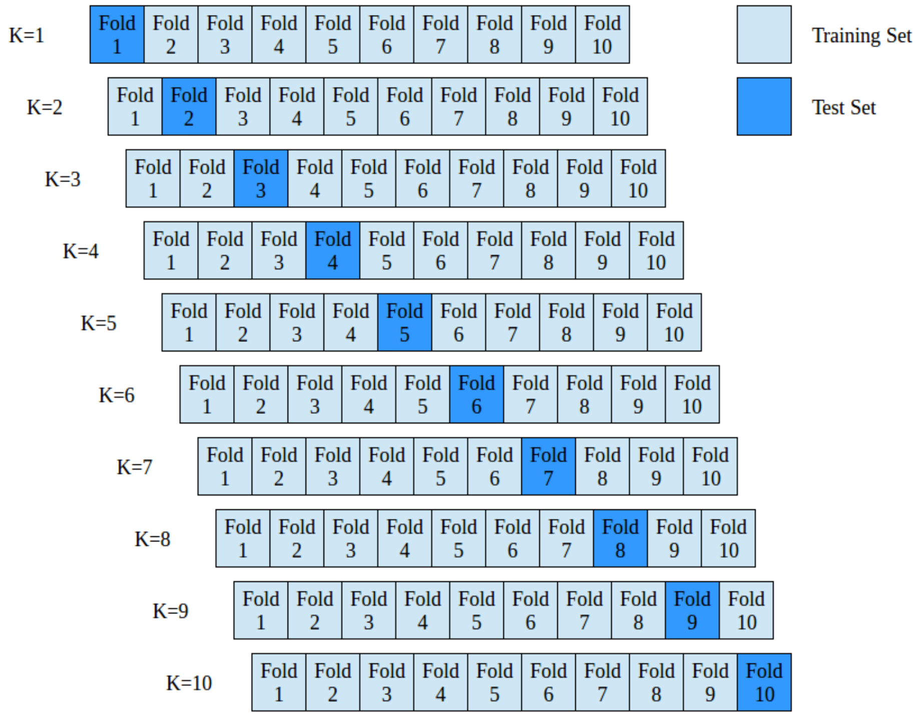

Figure 9. K-fold Cross-Validation Method

BIBLIOGRAPHY

Kamran, S. S., Haleem, A., Bahl, S., Javaid, M., Prakash, C., & Budhhi, D. (2022). Artificial intelligence and advanced materials in automotive industry: Potential applications and perspectives. Materials Today: Proceedings, 62, 4207-4214.

Fong, T., Thorpe, C., & Baur, C. (2001). Collaborative control: A robot-centric model for vehicle teleoperation (Vol. 1). Pittsburgh: Carnegie Mellon University, The Robotics Institute.

Hong, P., Turk, M., & Huang, T. S. (2000, March). Gesture modeling and recognition using finite state machines. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) (pp. 410-415). IEEE.

Lin, H. I., Hsu, M. H., & Chen, W. K. (2014, August). Human hand gesture recognition using a convolution neural network. In 2014 IEEE International Conference on Automation Science and Engineering (CASE) (pp. 1038-1043). IEEE.

Gao, Q., Liu, J., & Ju, Z. (2021). Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction. Expert Systems, 38(5), e12490.

Hakim, N. L., Shih, T. K., Kasthuri Arachchi, S. P., Aditya, W., Chen, Y. C., & Lin, C. Y. (2019). Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model. Sensors, 19(24), 5429.

Hong, P., Turk, M., & Huang, T. S. (2000, September). Constructing finite state machines for fast gesture recognition. In Proceedings 15th International

Conference on Pattern Recognition. ICPR-2000 (Vol. 3, pp. 691-694). IEEE.

Bedregal, B. C., Costa, A. C., & Dimuro, G. P. (2006). Fuzzy rule-based hand gesture recognition. In Artificial Intelligence in Theory and Practice: IFIP 19th World Computer Congress, TC 12: IFIP AI 2006 Stream, August 21–24, 2006, Santiago, Chile 1 (pp. 285-294). Springer US.

Verma, R., & Dev, A. (2009, October). Vision based hand gesture recognition using finite state machines and fuzzy logic. In 2009 International Conference on Ultra Modern Telecommunications & Workshops (pp. 1-6). IEEE.

Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008, December). A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In 2008 19th international conference on pattern recognition (pp. 1-4). IEEE.

Zeng, Z., Gong, Q., & Zhang, J. (2019). CNN Model Design of Gesture Recognition Based on Tensorflow Framework. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*.

Islam, M. Z., Hossain, M. S., ul Islam, R., & Andersson, K. (2019, May). Static hand gesture recognition using convolutional neural network with data augmentation. In 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 324-329). IEEE.

Oren, I. (2023, March 14). SAFe Team Kanban - Scaled Agile Framework. Scaled Agile Framework. https://scaledagileframework.com/safe-team-kanban/

McDonald, K. (2023, April 15). What is Kanban?. Agile Alliance. https://www.agilealliance.org/glossary/kanban/#q=~(infinite~false~filters~(postType~(~'page~'post~'aa_book~'aa_event_session~'aa_experience_report~'aa_glossary~'aa_research_paper~'aa_video)~tags~(~'kanban))~searchTerm~'~sort~false~sortDirection~'asc~page~1)

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.

Zafar, I., Tzanidou, G., Burton, R., Patel, N., & Araujo, L. (2018). Hands-On Convolutional Neural Networks with TensorFlow: Solve computer vision problems with modeling in TensorFlow and Python. Packt Publishing Ltd.

Manaswi, N. K., & Manaswi, N. K. (2018). Rnn and lstm. Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras, 115-126.

Yamashita, R., Nishio, M., G, R. K., DO, & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights Into Imaging, 9(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Johnston, M. V., & Bangalore, S. (2000). Finite-state multimodal parsing and understanding. https://doi.org/10.3115/990820.990874

Aksac, A., Ozturk, O., & Özyer, T. (2011). Real-time multi-objective hand posture/gesture recognition by using distance classifiers and finite state machine for virtual mouse operations. In International Conference on Electrical and Electronics Engineering. http://www.emo.org.tr/ekler/08ace10be6daff3_ek.pdf

Ben-Ari, M., & Mondada, F. (2017). Elements of Robotics. Springer.