**Controlling Robot Car Movement Using Hand Gesture Recognition**

**ABSTRACT**

This paper describes the implementation of movement control for robotic car using hand gesture recognition which uses deep learning algorithm. Therefore, proposed technique is hassle free as control is not based on joysticks or switches. There are six conditions considered for robot car movement control as ‘Backword’, ‘Forward’, ‘Left’, ‘No-Motion’, ‘Right’ and ‘Stop’ using different hand gestures. There are many researchers worked on this area using different sensors, machine learning algorithms and deep learning algorithms. Limitations of the state of art techniques are studied in this paper and designed a new modified convolutional neural network (CNN) for gesture recognition which controls movement of robot car. Dataset is created which generates 1000 Gray scale images for each type of gesture. Training modified CNN model gives prediction accuracy of 98.4024 % while random forest machine learning classifier gives prediction accuracy of 69%. It is observed that proposed model gives better accuracy compared to state of art technique for controlling movement of robot car using hand gesture. Obtained hand gesture class can be send to robot using Arduino controller for controlling movement.

***Keywords:*** Robot Car Movement, Gesture Recognition, Random Forest, Deep Learning, CNN, layer modification, Arduino-Uno Controller.

**CHAPTER 1**

**INTRODUCTION**

Controlling a robot car has many benefits in industries, for disabled person, in chemical laboratories, in defence, etc. There is huge demand of robot car based on automatic movement without use of switches and joystick.

Robot car can be controlled with hand gesture recognition and it has many applications such as manufacturing, medical, military, construction sectors. Hand gestures can be recognised to move the robotic car in five different directions such as stop, forward, backward, left and right.

Different services and operations can be handled to robot as it is intelligent machine. Productivity of the work is increased by using robot as time required to perform task is very less for robotic car. There basically two categories of robotic cars as robotic cars controlled by remote and second one is robotic car controlled autonomously. Robotic car controlled by remote includes robot controlled by gestures. Robots controlled autonomously include line and edge sensing robot [1].

Human feelings can be shared using either sign language or hand gestures. Without use of tele-operated robots and special hardware, with the help of gestures as well as sign, robotic movement can be controlled. With the help of temporal features hand or sign can be identified which then send to robot using controller [2].

Today’s world demands autonomous system which gives response very rapid without intervention of human being. Robotics field is growing very rapidly as per the technology demand causing harm to human. Robot need to be more precise to get real time environment. Mouse and keyboards are used in traditional methods for controlling robot. Image processing technique plays major role in hand gesture recognition-based robot control. Implemented algorithm should be invariant to factors like rotation, orientation, scaling, etc. Gaming field also uses gesture-based device control for interactive application [3].Introduction to robotic car movement based on hand gesture recognition is explained in chapter I, in chapter II , literature survey of existing works is explained. Chapter III , discuss about proposed methodology while chapter IV , describes results analysis and conclusion is explained at the end.

**What Is Machine Learning and How Does It Work?**

For starters, machine learning is a core sub-area of Artificial Intelligence (AI). ML applications learn from experience (or to be accurate, data) like humans do without direct programming. When exposed to new data, these applications learn, grow, change, and develop by themselves. In other words, machine learning involves computers finding insightful information without being told where to look. Instead, they do this by leveraging algorithms that learn from data in an iterative process.

The concept of machine learning has been around for a long time (think of the World War II Enigma Machine, for example). However, the idea of automating the application of complex mathematical calculations to big data has only been around for several years, though it’s now gaining more momentum. At a high level, machine learning is the ability to adapt to new data independently and through iterations. Applications learn from previous computations and transactions and use “pattern recognition” to produce reliable and informed results.

**How Does Machine Learning Work?**

Machine Learning is, undoubtedly, one of the most exciting subsets of Artificial Intelligence. It completes the task of learning from data with specific inputs to the machine. It’s important to understand what makes Machine Learning work and, thus, how it can be used in the future. The Machine Learning process starts with inputting training data into the selected algorithm.

Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily. New input data is fed into the machine learning algorithm to test whether the algorithm works correctly. The prediction and results are then checked against each other.

A robot is an intelligent machine that is commanded by a computer application to perform various operations and services. Robots play an essential role in automation across all sectors like manufacturing, military, construction, and medical [1] etc. Robots not only helps humans to save time but also use to increases the productivity [2], efﬁciency, reliability, reduces the use of resources to save energy, and reduce the running cost etc.

Meanwhile, robots play a significant role in providing help in such tasks that cannot be done smoothly by disabled persons, i.e., controlling the car by physical devices has become very successful. There are two categories of robots such as: autonomous robots (edge sensing robots [3], line sensing [4]) and remote-controlled robots (gesture-controlled robots [5]). Therefore, the employment of gesture-controlled robots is one of the most elegant and aesthetic term to catch the human-gesture that is difficult to understand by machine.

There are many methods to capture gesture commonly using data glove [6], camera [7], infrared waves [8], tactile [9], acoustic [10] and with motion technological means [11]. These embedded systems [6–11] are designed for particular control and can be optimized to increase reliability, performance and reduce the cost and size of the device. Moreover, researchers are showing tremendous interest in gesture recognition, building robots and many other devices that are directly controlled by human gestures.

Gesture control mechanism is applied in various fields like socially assistive robots, augmented reality, emotion detection from facial expressions, recognition of sign languages [12–14] etc. Furthermore, the emotional gesture [12] identification from the face is also been investigated. Similarly, with the arrival of smartphone and other advanced technologies, ntial for connecting robot with the internet to allow users to control robot from anywhere and anytime.

These wireless systems are providing vital help to robot self-regulation systems by using Wi-Fi and cloud computing, etc. As far as we know, a need still exists for the design of an automated robot control system that supports hand-gesture recognition, android mobile application control and voice recognition concepts, while monitoring the obstacles. In this paper, we introduce the design and experimentally demonstrate a robot-car that can be controlled with hand movement by using the technique of hand-gesture.

This work is accomplished with the conventional arrangements of the Arduino microcontroller, accelerometer, RF transmitter / receiver, Android mobile-application, Bluetooth and motor module. The proposed robot-car is controlled via gesture recognition technique by measuring the angles and position of the hand. In this way, the robot will move according to the movement of the hand. Furthermore, we extend this system to allow the robot to be controlled by just a click on the cellphone with an Android operating system, and the voice recognition via Bluetooth technology. There are two main controlling ways in android application.

First one is the touch buttons, the robot will move accordingly as the user touches the button. The second one is voice recognition; the robot will follow and move accordingly as the user says the operating command. Most importantly, an obstacle detection sensor is set to detect the obstacle in front of it, and when sensor detects the obstacle, it stops moving. Hence, the proposed systems of controlling the robot car with both gesture control and android application control are performed and displayed at lab-scale prototype to confirm that the proposed designs can be easily implemented in the large and real-scale conditions in future.

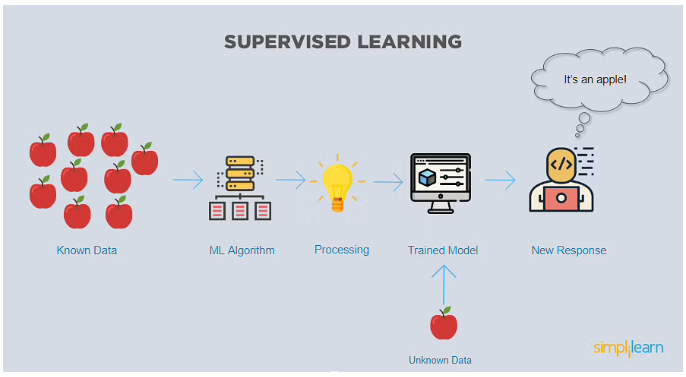
The remaining content of the paper is ordered as follows. In Section 2, the idea of the automatic control robot car is introduced with a detailed explanation of the electronic components that are used in the proposed system, based on the gesture recognition and with the mobile-based android application. Also, the experimental results of proposed systems at lab-scale prototype are presented in Section 3, and Section 4 concludes the paper.

**What are the Different Types of Machine Learning?**

Machine Learning is complex, which is why it has been divided into two primary areas, supervised learning and unsupervised learning. Each one has a specific purpose and action, yielding results and utilizing various forms of data. Approximately 70 percent of machine learning is supervised learning, while unsupervised learning accounts for anywhere from 10 to 20 percent. The remainder is taken up by reinforcement learning.

**1. Supervised Learning**

In supervised learning, we use known or labeled data for the training data. Since the data is known, the learning is, therefore, supervised, i.e., directed into successful execution. The input data goes through the Machine Learning algorithm and is used to train the model. Once the model is trained based on the known data, you can use unknown data into the model and get a new response.



In this case, the model tries to figure out whether the data is an apple or another fruit. Once the model has been trained well, it will identify that the data is an apple and give the desired response.

Here is the list of top algorithms currently being used for supervised learning are:

1. Polynomial regression

2. Random forest

3. Linear regression

4. Logistic regression

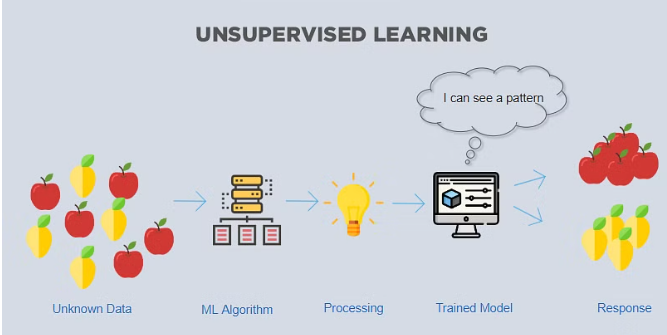
5. Decision trees

6. K-nearest neighbors

7. Naive Bayes

**2. Unsupervised Learning**

In unsupervised learning, the training data is unknown and unlabeled - meaning that no one has looked at the data before. Without the aspect of known data, the input cannot be guided to the algorithm, which is where the unsupervised term originates from. This data is fed to the Machine Learning algorithm and is used to train the model. The trained model tries to search for a pattern and give the desired response. In this case, it is often like the algorithm is trying to break code like the Enigma machine but without the human mind directly involved but rather a machine.



In this case, the unknown data consists of apples and pears which look similar to each other. The trained model tries to put them all together so that you get the same things in similar groups.

The top 7 algorithms currently being used for unsupervised learning are:

1. Partial least squares

2. Fuzzy means

3. Singular value decomposition

4. K-means clustering

5. Apriori

6. Hierarchical clustering

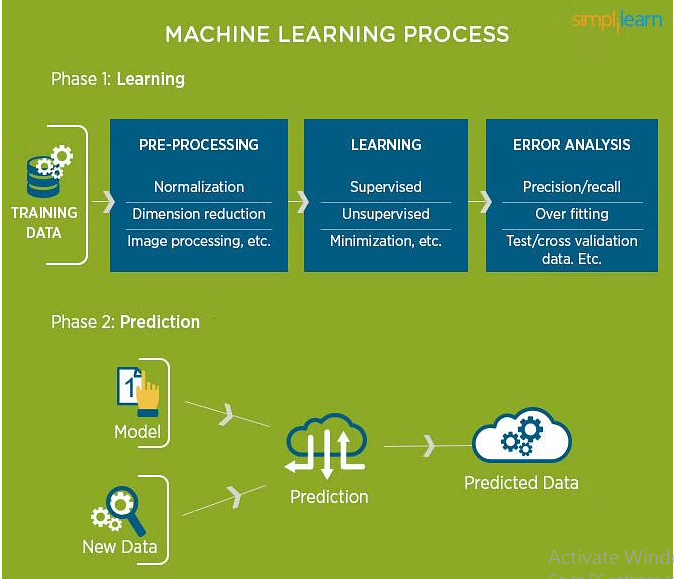
7. Principal component analysis

**3. Reinforcement Learning**

Like traditional types of data analysis, here, the algorithm discovers data through a process of trial and error and then decides what action results in higher rewards. Three major components make up reinforcement learning: the agent, the environment, and the actions. The agent is the learner or decision-maker, the environment includes everything that the agent interacts with, and the actions are what the agent does.

**Why is Machine Learning Important?**

To better answer the question :what is machine learning” and understand the uses of Machine Learning, consider some of the applications of Machine Learning: the self-driving Google car, cyber fraud detection, and online recommendation engines from Facebook, Netflix, and Amazon. Machines make all these things possible by filtering useful pieces of information and piecing them together based on patterns to get accurate results.



**Main Uses of Machine Learning**

Typical results from machine learning applications usually include web search results, real-time ads on web pages and mobile devices, email spam filtering, network intrusion detection, and pattern and image recognition. All these are the by-products of using machine learning to analyze massive volumes of data.

Since the introduction of the most common input computer devices not a lot have changed. This is probably because the existing devices are adequate. It is also now that computers have been so tightly integrated with everyday life, that new applications and hardware are constantly introduced. The means of communicating with computers at the moment are limited to keyboards, mice, light pen, trackball, keypads etc. These devices have grown to be familiar but inherently limit the speed and naturalness with which we interact with the computer. As the computer industry follows Moore’s Law since middle 1960s, powerful machines

are built equipped with more peripherals. Vision based interfaces are feasible and at the present moment the computer is able to “see”. Hence users are allowed for richer and user-friendlier man-machine interaction. This can lead to new interfaces that will allow the deployment of new commands that are not possible with the current input devices. Plenty of time will be saved as well. Recently, there has been a surge in interest in recognizing human hand gestures. Handgesture recognition has various applications like computer games, machinery control (e.g. crane), and thorough mouse replacement. One of the most structured sets of gestures belongs to sign language. In sign language, each gesture has an assigned meaning (or

meanings). Computer recognition of hand gestures may provide a more natural-computer interface, allowing people to point, or rotate a CAD model by rotating their hands. Hand gestures can be classified in two categories: static and dynamic. A static gesture is a particular hand configuration and pose, represented by a single image. A dynamic gesture is a moving gesture, represented by a sequence of images. We will focus on the recognition of static images. Interactive applications pose particular challenges. The response time should be very fast. The user should sense no appreciable delay between when he or she makes a gesture or

**Introduction**

motion and when the computer responds. The computer vision algorithms should be reliable and work for different people. There are also economic constraints: the vision-based interfaces will be replacing existing ones, which are often very low cost. A hand-held video game controller and a television remote control each cost about $40. Even for added functionality, consumers may not want to spend more. When additional hardware is needed the cost is considerable higher. Academic and industrial researchers have recently been focusing on analyzing images of people. While researchers are making progress, the problem is hard and many present day

algorithms are complex, slow or unreliable. The algorithms that do run near real-time do so on computers that are very expensive relative to the existing hand-held interface devices.

**Applications:-**

Creating a proper sign language (ASL – American Sign Language at this case) dictionary is not the desired result at this point. This would combine advanced grammar and syntax structure understanding of the system, which is outside the scope of this project. The American Sign Language will be used as the database since it’s a tightly structured set. From that point further applications can be suited. The distant (or near?) future of computer interfaces could have the usual input devices and in conjunction with gesture recognition some of the user’s feelings would be perceived as well. Taking ASL recognition further a full real-time dictionary could be created with the use of video. As mentioned before this would require some Artificial Intelligence for grammar and syntax purposes. Another application is huge database annotation. It is far more efficient when properly executed by a computer, than by a human.

In many ways, sign languages are like spoken languages: they are natural

languages that arise spontaneously wherever there is a community of communicators; they effectively fullfill all the social and mental functions of spoken languages; and they’re acquired without instruction by children, given normal exposure and interaction. These characteristics have led many linguists to expect sign languages to be similar to spoken languages in significant ways. But sign languages are different too: as manual visual languages, sign languages exploit a completely different physical medium from the vocal-auditory system of spoken languages. These two dramatically different physical modalities are also likely to have an effect on the structure of the languages through which they are transmitted. It is of special interest, then, to compare natural languages in the two modalities. Where the two systems converge, universal linguistic properties are revealed. Where they diverge, the physical medium of transmission is implicated, and its contribution to the form of language in both modalities illuminated. Neither can be seen quite so clearly if linguists restrict their study to spoken language alone (or to sign language alone). For this and other related reasons, it is often remarked that sign languages provide us with a natural laboratory for studying the basic characteristics of all human language. Once the existence of natural language in a second modality is acknowledged, questions like the following arise: How are such languages born? Are the central linguistic properties of sign languages parallel to those of spoken languages? Is sign language acquired by children in the same stages and time frame in which hearing children acquire spoken language? Are the same areas of the brain responsible for language in both modalities? What role does modality play in structuring language? In other words, within the architecture of human cognition, do we find the structure of one language ‘faculty’ or two? While there is no conclusive answer to this deceptively simple question, an impressive body of research has greatly expanded our understanding of the

issues underlying it.

Evolution made language possible scores of millennia ago, and there is no human community without it. What sign language teaches us is that humans have a natural propensity for language in two different modalities: vocal-auditory and manual-visual. Since the human ability to use language is so old, and since speech is the predominant medium for its transmission, it seems that spoken languages themselves are either also very old or are descended from other languages with a long history. But sign languages do not have the same histories as spoken languages because special conditions are required for them to arise and persevere, and for this reason they can offer unique insight into essential features of human language. The first lesson sign language teaches us is that, given a community of humans, language inevitably emerges. While we have no direct evidence of the emergence of any spoken language, we can get much closer to the origin of a sign language, and, in rare instances, even watch it come into being. Wherever deaf people have an opportunity to gather and interact regularly, a sign language is born. Typically, deaf people make up a very small percentage of the population (about 0.23% in the United States, according to the National Center for Health

Statistics, 1994), so that in any given local social group, there may be no deaf people at all or very few of them. The most common setting in which a deaf community can form, then, is a school for deaf children. Such schools only began to be established about two hundred years ago in Europe and North America. On the basis of this historical information and some reported earlier observations of groups of people using sign language, it is assumed that the oldest extant sign languages do not date back farther than about 300 years (Woll et al, 2001). Currently, linguists have the rare opportunity to observe the emergence and development of a sign language from the beginning in a school established in Nicaragua only about 25 years ago, an opportunity that is yielding

very interesting results. Graduates of such schools sometimes choose to concentrate in certain urban areas, and wider communities arise and grow, creating their own social networks, institutions, and art forms, such as visual poetry (Sutton Spence and Woll, 1999; Padden and Humphries, in press; Sandler and Lillo-Martin, in press). Deaf society is highly developed in some places, and the term ‘Deaf’ with a capital D has come to refer to members of a minority community with its own language and culture, rather than to people with an auditory deficit.

It is not only the genesis of a sign language that is special; the way in which it is

passed down from generation to generation is unusual as well. Typically, fewer than 10% of deaf children acquire sign language from deaf parents, and of those deaf parents, only a small percentage are themselves native signers. The other 90+% of deaf children have hearing parents and may only be exposed to a full sign language model when they get to school. These social conditions taken together with certain structural properties of sign languages have prompted some linguists to compare them to spoken creoles (Fischer, 1978). Another way in which a deaf social group and concomitant sign language can form is through the propagation of a genetic trait within a small village or town through consanguineous marriage, resulting in a proportionately high incidence of deafness, and the spread of the sign language among both deaf and hearing people. Potentially, this kind of situation can allow us to observe the genesis and development of a language in a natural community setting. Though the existence of such communities has been reported here and there (see especially Groce, 1985), no detailed linguistic study of a language arising in such a community has yet been provided. These, then, are the ways in which sign languages happen. The existence of many sign languages around the world -- the number 103 found in the Ethnologue database is probably an underestimate – confirms the claim that the emergence of a highly structured communication system among humans is inevitable. If the oral-aural channel is unavailable, language springs forth in the manual-visual modality.

Not only does such a system emerge in the absence of audition, but its kernels can

be observed even in the absence of both a community and a language model. Deaf

children who live in hearing households where only oral language is used, who have not yet experienced speech training, and thus have no accessible language model, devise their own systematic means of communication called home sign, studied in exquisite detail by Goldin-Meadow and her colleagues (Goldin-Meadow, 2003). The gesture talk of these children contains the unmistakable imprint of a real linguistic system, and as such it offers a unique display of the fundamental human genius for language. At the same time, the form and content of home sign are rudimentary, and do not approach the richness and complexity of a language used by a community, spoken or signed. This confronts us with another important piece of information: language as we know it is a social phenomenon. Although each brain possesses the potential for language, it takes more than one brain to create a complex linguistic system.

**HUMAN COMPUTER INTERFACE SYSTEM**

Computer is used by many people either at their work or in their spare-time. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse. Every new device can be seen as an attempt to make the computer more intelligent and making humans able to perform more complicated communication with the computer. This has been possible due to the result oriented efforts made by computer professionals for creating successful human computer interfaces. As the complexities of human needs have turned into many folds and continues to grow so, the need for Complex programming ability and intuitiveness are critical attributes of computer programmers to survive in a competitive environment. The computer programmers have been incredibly successful in easing the communication between computers and human. With the emergence of every new product in the market; it attempts to ease the complexity of jobs performed. For instance, it has helped in facilitating tele operating, robotic use, better human control over complex work systems like cars, planes and monitoring systems. Earlier, Computer programmers were avoiding such kind of complex programs as the focus was more on speed than other modifiable features. However, a shift towards a user friendly environment has driven them to revisit the focus area.

The idea is to make computers understand human language and develop a user friendly human computer interfaces (HCI). Making a computer understand speech, facial expressions and human gestures are some steps towards it. Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time. Since human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. The project aims to determine human gestures by creating an HCI. Coding of these gestures into machine language demands a complex programming algorithm. An overview of gesture recognition system is given to gain knowledge.

**GESTURES:-**

It is hard to settle on a specific useful definition of gestures due to its wide variety of applications and a statement can only specify a particular domain of gestures. Many researchers had tried to define gestures but their actual meaning is still arbitrary. Bobick and Wilson have defined gestures as the motion of the body that is intended to communicate with other agents. For a successful communication, a sender and a receiver must have the same set of information for a particular gesture. As per the context of the project, gesture is defined as an expressive movement of body parts which has a particular message, to be communicated precisely between a sender and a receiver. A gesture is scientifically categorized into two distinctive categories: dynamic and static. A dynamic gesture is intended to change over a period of time whereas a static gesture is observed at the spurt of time. A waving hand means goodbye is an example of dynamic gesture and the stop sign is an example of static gesture. To understand a full message, it is necessary to interpret all the static and dynamic gestures over a period of time. This complex process is called gesture recognition. Gesture recognition is the process of recognizing and interpreting a stream continuous sequential gesture from the given set of input data.

**GESTURE MODELING**

In order to systematically discuss the literature on gesture interpretation, it is important to first consider what model the authors have used for the hand gesture. In fact, the scope of a gestural interface for HCI is directly related to the proper modeling of hand gestures. How to model hand gestures depends primarily on the intended application within the HCI context. For a given application, a very

coarse and simple model may be sufficient. However, if the purpose is a natural-like interaction, a model has to be established that allows many if not all natural gestures to be interpreted by the computer. The following discussion addresses

the question of modeling of hand gestures for HCI.

**Definition of Gestures**

Outside the HCI framework, hand gestures cannot be easily defined. The definitions, if they exist, are particularly related to the communicational aspect of the human hand and body movements. Webster’s Dictionary, for example,

defines gestures as ”...the use of motions of the limbs or body as a means of expression; a movement usually of the body or limbs that expresses or emphasizes an idea, sentiment, or attitude.” Psychological and social studies tend to narrow this broad definition and relate it even more to man’s expression and social interaction . However, in the domain of HCI the notion of gestures is somewhat different. In a computer controlled environment one wants to use the human hand to perform tasks that mimic both the natural use of the hand as a manipulator, and its use in human-machine communication (control of computer/ machine functions through gestures). Classical definitions of gestures, on the other hand, are rarely, if ever, concerned with the former mentioned use of the human hand (so called practical gestures ).

Hand gestures are a means of communication, similar to spoken language. The production and perception of gestures can thus be described using a model commonly found in the field of spoken language recognition [85], [100]. An

interpretation of this model, applied to gestures, is depicted in Fig. 2. According to the model, gestures originate as a gesturer’s mental concept, possibly in conjunction with speech. They are expressed through the motion of arms and hands, the same way speech is produced by air stream modulation through the human vocal tract. Also, observers perceive gestures as streams of visual images which they interpret using the knowledge they possess about those

gestures. The production and perception model of gestures can also be summarized in the following form:

Transformations *T*. can be viewed as different *models*: *Thg* is a model of hand or arm motion given gesture *G*, *Tvh* is a model of visual images given hand or arm motion *H*, and *Tvg* describes how visual images *V* are formed given some gesture *G*. The models are parametric, with the parameters belonging to their respective parameter spaces 0 7.. In light of this notation, one can say that the aim of visual interpretation of hand gestures is to infer gestures *G* from their visual

images *V* using a suitable gesture model *Tvg*, or In the context of visual interpretation of gestures, it may then be useful to consider the following definition of gestures:

**GESTURE BASED APPLICATIONS:-**

Gesture based applications are broadly classified into two groups on the basis of their purpose: multidirectional control and a symbolic language.

***3D Design***: CAD (computer aided design) is an HCI which provides a platform for interpretation and manipulation of 3-Dimensional inputs which can be the gestures. Manipulating 3D inputs with a mouse is a time consuming task as the task involves a complicated process of decomposing a six degree freedom task into at least three sequential two degree tasks. Massachuchetttes institute of technology has come up with the 3DRAW technology that uses a pen embedded in polhemus device to track the pen position and orientation in 3D.A 3space sensor is embedded in a flat palette, representing the plane in which the objects rest .The CAD model is moved synchronously with the users gesture movements and objects can thus be rotated and translated in order to view them from all sides as they are being created and altered.

***Tele presence***: There may raise the need of manual operations in some cases such as system failure or emergency hostile conditions or inaccessible remote areas. Often it is impossible for human operators to be physically present near the machines. Tele presence is that area of technical intelligence which aims to provide physical operation support that maps the operator

arm to the robotic arm to carry out the necessary task, for instance the real time ROBOGEST system constructed at University of California, San Diego presents a natural way of controlling an outdoor autonomous vehicle by use of a language of hand gestures . The prospects of tele presence includes space, undersea mission, medicine manufacturing and in maintenance of nuclear power reactors.

***Virtual reality***: Virtual reality is applied to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds. Most current virtual reality environments are primarily visual experiences, displayed either on a computer screen or through special stereoscopic displays [6]. There are also some simulations include additional sensory information, such as sound through speakers or headphones. Some advanced, haptic systems now include tactile information, generally known as force feedback, in medical and gaming applications.

***Sign Language***: Sign languages are the most raw and natural form of languages could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. It has started even before the emergence of spoken languages. Since then the sign language has evolved and been adopted as an integral part of our day to day communication process. Now, sign languages are being used extensively in international sign use of deaf and dumb, in the world of sports, for religious practices and also at work places . Gestures are one of the first forms of communication when a child learns to express its need for food, warmth and comfort. It enhances the emphasis of spoken language and helps in expressing thoughts and feelings effectively.

A simple gesture with one hand has the same meaning all over the world and means either ’hi’ or ‘goodbye’. Many people travel to foreign countries without knowing the official language of the visited country and still manage to perform communication using gestures and sign language. These examples show that gestures can be considered international and used almost all over the world. In a number of jobs around the world gestures are means of communication.

In airports, a predefined set of gestures makes people on the ground able to communicate with the pilots and thereby give directions to the pilots of how to get off and on the run-way and the referee in almost any sport uses gestures to communicate his decisions. In the world of sports gestures are common. The pitcher in baseball receives a series of gestures from the coach to help him in deciding the type of throw he is about to give. Hearing impaired people have over the years developed a gestural language where all defined gestures have an assigned meaning. The language allows them to communicate with each other and the world they live in.

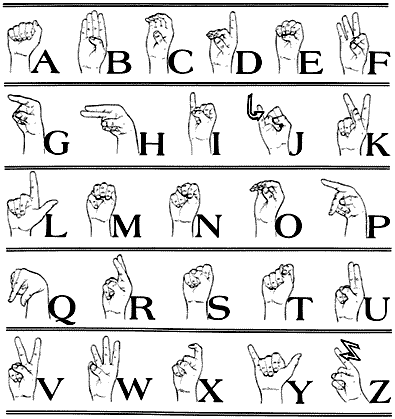


Fig 1.1 American Sign Language

The recognition of gestures representing words and sentences as they do in American and Danish sign language undoubtedly represents the most difficult recognition problem of those applications mentioned before. A functioning sign language recognition system could provide an opportunity for the deaf to communicate with non-signing people without the need for an interpreter. It could be used to generate speech or text making the deaf more independent. Unfortunately there has not been any system with these capabilities so far. In this project our aim is to develop a system which can classify sign language accurately.

**CHAPTER 2**

**LITERATURE SURVEY**

**[1] Saleem Ullah1, Zain Mumtaz1,etAl. (2019) ,An Automated Robot-Car Control System with Hand-Gestures and Mobile Application Using Arduino, © 2019 by the author(s). Distributed under a Creative Commons CC BY license,** <https://www.researchgate.net/publication/330182990>

Gesture recognition has always been a technique to decrease the distance between the physical and the digital world. In this work, we introduce an Arduino based vehicle system which no longer requires manual controlling of the cars. The proposed work is achieved by utilizing the Arduino microcontroller, accelerometer, RF sender/receiver, and Bluetooth. Two main contributions are presented in this work. Firstly, we show that the car can be controlled with hand-gestures according to the movement and position of the hand.

Secondly, the proposed car system is further extended to be controlled by an android based mobile application having different modes (e.g., touch buttons mode, voice recognition mode). In addition, an automatic obstacle detection system is introduced to improve the safety measurements to avoid any hazards. The proposed systems are designed at lab-scale prototype to experimentally validate the efﬁciency, accuracy, and affordability of the systems. We remark that the proposed systems can be implemented under real conditions at large-scale in the future that will be useful in automobiles and robotics applications.

**[1] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, and F. A. Gonzalez**. An unsupervised feature learning framework for ´ basal cell carcinoma image analysis. Artificial intelligence in medicine, 2015.

The paper addresses the problem of automatic detection of basal cell carcinoma (BCC) in histopathology images. In particular, it proposes a framework to both, learn the image representation in an unsupervised way and visualize discriminative features supported by the learned model.

The proposed UFL-representation-based approach outperforms state-of-the-art methods for BCC detection. Thanks to its visual interpretation layer, the method is able to highlight discriminative tissue regions providing a better diagnosis support. Among the different UFL strategies tested, TICA-learned features exhibited the best performance thanks to its ability to capture low-level invariances, which are inherent to the nature of the problem.

**[2] J. Arroyo and B. Zapirain**. Automated detection of melanoma in dermoscopic images. In J. Scharcanski and M. E. Celebi, editors, Computer Vision Techniques for the Diagnosis of Skin Cancer, Series in BioEngineering, pages 139–192. Springer Berlin Heidelberg, 2014.

The incidence of malignant melanoma continues to increase worldwide. This cancer can strike at any age; it is one of the leading causes of loss of life in young persons. Since this cancer is visible on the skin, it is potentially detectable at a very early stage when it is curable. New developments have converged to make fully automatic early melanoma detection a real possibility. First, the advent of dermoscopy has enabled a dramatic boost in clinical diagnostic ability to the point that melanoma can be detected in the clinic at the very earliest stages. The global adoption of this technology has allowed accumulation of large collections of dermoscopy images of melanomas and benign lesions validated by histopathology. The development of advanced technologies in the areas of image processing and machine learning have given us the ability to allow distinction of malignant melanoma from the many benign mimics that require no biopsy.

These new technologies should allow not only earlier detection of melanoma, but also reduction of the large number of needless and costly biopsy procedures. Although some of the new systems reported for these technologies have shown promise in preliminary trials, widespread implementation must await further technical progress in accuracy and reproducibility. In this paper, we provide an overview of computerized detection of melanoma in dermoscopy images. First, we discuss the various aspects of lesion segmentation. Then, we provide a brief overview of clinical feature segmentation. Finally, we discuss the classification stage where machine learning algorithms are applied to the attributes generated from the segmented features to predict the existence of melanoma.

The early detection of melanoma is essential for successful treatment. Because dermoscopy images are so inexpensive to obtain and so widely available, they provide the most viable option for application of new image processing and machine learning algorithms. Therefore, melanoma detection using dermoscopy images has the most potential for disruption of the current clinical paradigm of waiting until the melanoma is at a later stage and performing an excessive number of biopsies. The advent of a fast, accurate and cost-effective on-the-spot technology, in the clinic or even at home, is most likely to be afforded by the type of computer analysis of dermoscopy images described here. Dermoscopy images come with various aberrations and artifacts and hence it is crucial to follow the proper steps and methods described here to remedy these abnormalities and achieve a correct diagnosis. Lesion segmentation with acceptable tolerance allows for acceptable precision in feature segmentation which in turn helps in maximizing classification accuracy. Although lesion segmentation, feature segmentation, feature generation and classification are the major steps, proper attention should be given to the auxiliary steps which in most cases are the major contributors to an exemplary outcome.

**[3] C. Barata, J. Marques, and T. Mendonc¸a**. Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors. In M. Kameland A. Campilho, editors, Image Analysis and Recognition, volume 7950 of Lecture Notes in Computer Science, pages 547–555. Springer Berlin Heidelberg, 2013.

The identification of melanomas in dermoscopy images is still an up to date challenge. Several Computer Aided-Diagnosis Systems for the early diagnosis of melanomas have been proposed in the last two decades. This chapter presents an approach to diagnose melanomas using Bag-of-features, a classification method based on a local description of the image in small patches. Moreover, a comparison between color and texture descriptors is performed in order to assess their discriminative power. The presented results show that local descriptors allow an accurate representation of dermoscopy images and achieve good classification scores: Sensitivity = 93% and Specificity = 88%. Furthermore it shows that color descriptors perform better than texture ones in the detection of melanomas.

This chapter investigates the applicability of local color and texture features to the melanoma classification problem. Several factors associated with the performance of BoF were tested, namely the type of descriptors used and the classification algorithm. The results show that individually color descriptors perform better than texture descriptors and that good classification results can be achieved using kNN (SE = 93%, SP = 85% with hLa∗b∗ and SE = 100%, SP = 75% with hL∗uv) and SVM (SE = 93%, SP = 88% with MOpp). The fusion of color and texture descriptors also achieved good results, with a score of SE = 96%, SP = 82% for the combination of Opp moments with Gabor and Laws texture descriptors. A simple analysis of the visual words showed that the dictionary obtained using BoF has potential to be used as a detector/identifier for specific dermoscopic features and patterns. Future work will rely on testing this hypothesis in order to develop a more medical oriented system. Moreover, sparse sampling methods should be tested in order to compare their performances with that of the dense sampling used in this chapter. Finally, high-level descriptors should be tested as well.

**[4] Y. Bengio, A. Courville, and P. Vincent**. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell., 35(8):1798–1828, Aug. 2013.

The success of machine learning algorithms generally depends on data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different explanatory factors of variation behind the data. Although specific domain knowledge can be used to help design representations, learning with generic priors can also be used, and the quest for AI is motivating the design of more powerful representation-learning algorithms implementing such priors. This paper reviews recent work in the area of unsupervised feature learning and deep learning, covering advances in probabilistic models, auto-encoders, manifold learning, and deep networks. This motivates longer-term unanswered questions about the appropriate objectives for learning good representations, for computing representations (i.e., inference), and the geometrical connections between representation learning, density estimation and manifold learning.

This review of representation learning and deep learning has covered three major and apparently disconnected approaches: the probabilistic models (both the directed kind such as sparse coding and the undirected kind such as Boltzmann machines), the reconstruction-based algorithms related to auto encoders, and the geometrically motivated manifold-learning approaches. Drawing connections between these approaches is currently a very active area of research and is likely to continue to produce models and methods that take advantage of the relative strengths of each paradigm.

**[5] H. Chang, Y. Zhou, A. Borowsky, K. Barner, P. Spellman, and B. Parvin**. Stacked predictive sparse decomposition for classification of histology sections. International Journal of Computer Vision, 113(1):3–18, 2014.

Image-based classification of histology sections, in terms of distinct components (e.g., tumor, stroma, normal), provides a series of indices for histology composition (e.g., the percentage of each distinct components in histology sections), and enables the study of nuclear properties within each component. Furthermore, the study of these indices, constructed from each whole slide image in a large cohort, has the potential to provide predictive models of clinical outcome. For example, correlations can be established between the constructed indices and the patients’ survival information at cohort level, which is a fundamental step towards personalized medicine. However, performance of the existing techniques is hindered as a result of large technical variations (e.g., variations of color/textures in tissue images due to non-standard experimental protocols) and biological heterogeneities (e.g., cell type, cell state) that are always present in a large cohort.

We propose a system that automatically learns a series of dictionary elements for representing the underlying spatial distribution using stacked predictive sparse decomposition. The learned representation is then fed into the spatial pyramid matching framework with a linear support vector machine classifier. The system has been evaluated for classification of distinct histological components for two cohorts of tumor types. Throughput has been increased by using of graphical processing unit (GPU), and evaluation indicates a superior performance results, compared with previous research.

**[6] N. Cox and I. Coulson**. Diagnosis of skin disease. Rook’s Textbook of Dermatology, 7th edn. Oxford: Blackwell Science, 5, 2004.

The late Arthur Rook established the Textbook of Dermatology as the most comprehensive work of reference available to the dermatologist. Covering all aspects of skin disease from basic science through pathology and epidemiology to clinical practice, the text is recognized for its unparalleled coverage of diagnosis. Hailed by reviewers as 'a thorough, modern masterpiece' and 'the best textbook of dermatology in the world', and trusted by dermatologists around the world for accurate and comprehensive coverage, this clinical classic is the definitive source of information for all dermatologists. The new edition of this venerable classic extends the standard of excellence to include: All-new coverage of cosmetic dermatology and sexually transmitted diseases. More material on evidence-based dermatology. Increased coverage of dermoscopy. More emphasis on therapeutics throughout the set. More contributions from a greater variety of international experts. New page design with larger illustrations for more immediate recognition.

The 8th Edition marks the debut of the online edition of Rook's Textbook of Dermatology, allowing users the fastest possible access to the full range of knowledge on all known dermatological conditions. With fully searchable text and a fully searchable bank of more than 3,300 downloadable images, this online version puts specific information at your fingertips - when and where you need it - and is free with purchase of the four-volume set. The person-specific access code travels with you, not your computer, so you can check with Rook from any location. With the online version, you can: Search across all four volumes simultaneously. Search all images separately. Download images into presentations. Link directly to references via a range of sources. Rook's Textbook of Dermatology, in print and now online, provides a reliable, constant companion for all dermatologists.

**[7] A. Cruz-Roa, A. Basavanhally, F. Gonzalez, H. Gilmore, ´ M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi**. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In SPIE Medical Imaging, pages 904103–904103. International Society for Optics and Photonics, 2014.

This paper presents a deep learning approach for automatic detection and visual analysis of invasive ductal carcinoma (IDC) tissue regions in whole slide images (WSI) of breast cancer (BCa). Deep learning approaches are learn-from-data methods involving computational modeling of the learning process. This approach is similar to how human brain works using different interpretation levels or layers of most representative and useful features resulting into a hierarchical learned representation. These methods have been shown to outpace traditional approaches of most challenging problems in several areas such as speech recognition and object detection. Invasive breast cancer detection is a time consuming and challenging task primarily because it involves a pathologist scanning large swathes of benign regions to ultimately identify the areas of malignancy. Precise delineation of IDC in WSI is crucial to the subsequent estimation of grading tumor aggressiveness and predicting patient outcome. DL approaches are particularly adept at handling these types of problems, especially if a large number of samples are available for training, which would also ensure the generalizability of the learned features and classifier.

The DL framework in this paper extends a number of convolutional neural networks (CNN) for visual semantic analysis of tumor regions for diagnosis support. The CNN is trained over a large amount of image patches (tissue regions) from WSI to learn a hierarchical part-based representation. The method was evaluated over a WSI dataset from 162 patients diagnosed with IDC. 113 slides were selected for training and 49 slides were held out for independent testing. Ground truth for quantitative evaluation was provided via expert delineation of the region of cancer by an expert pathologist on the digitized slides. The experimental evaluation was designed to measure classifier accuracy in detecting IDC tissue regions in WSI. Our method yielded the best quantitative results for automatic detection of IDC regions in WSI in terms of F-measure and balanced accuracy (71.80%, 84.23%), in comparison with an approach using handcrafted image features (color, texture and edges, nuclear textural and architecture), and a machine learning classifier for invasive tumor classification using a Random Forest.

The best performing handcrafted features were fuzzy color histogram (67.53%, 78.74%) and RGB histogram (66.64%, 77.24%). Our results also suggest that at least some of the tissue classification mistakes (false positives and false negatives) were less due to any fundamental problems associated with the approach, than the inherent limitations in obtaining a very highly granular annotation of the diseased area of interest by an expert pathologist.

**[8] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio**. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pages 403– 410. Springer, 2013.

Skin diseases are very common in people’s daily life. Each year, millions of people in American are affected by all kinds of skin disorders. Diagnosis of skin diseases sometimes requires a high-level of expertise due to the variety of their visual aspects. As human judgment are often subjective and hardly reproducible, to achieve a more objective and reliable diagnosis, a computer aided diagnostic system should be considered. In this paper, we investigate the feasibility of constructing a universal skin disease diagnosis system using deep convolutional neural network (CNN). We train the CNN architecture using the 23,000 skin disease images from the Dermnet dataset and test its performance with both the Dermnet and OLE, another skin disease dataset, images. Our system can achieve as high as 73.1% Top-1 accuracy and 91.0% Top-5 accuracy when testing on the Dermnet dataset. For the test on the OLE dataset, Top-1 and Top-5 accuracies are 31.1% and 69.5%. We show that these accuracies can be further improved if more training images are used.

We have investigated the feasibility of building an universal skin disease classification system using deep CNN. We tackle this problem by fine-tuning ImageNetpretrained models (VGG16, VGG19, GoogleNet) with the Dermnet dataset. Our experiments show that the current state-ofart CNN models can achieve as high as 73.1% Top-1 ac- curacy (VGG19) and 91.0% Top-5 accuracy (GoogleNet) when testing on the Dermnet dataset. We further discover the performance of the CNN architecture when testing on a different dataset (OLE). We find the classification system can only achieve 24.8% Top-1 accuracy and 61.7% Top-5 accuracy due to the lack of a broader variance in the training set. We show that by increasing the variance of the training set the Top-1 and Top-5 accuracies can be improved to 31.1% and 69.5%.

**[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei.**Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255, June 2009.

The explosion of image data on the Internet has the potential to foster more sophisticated and robust models and algorithms to index, retrieve, organize and interact with images and multimedia data. But exactly how such data can be harnessed and organized remains a critical problem. We introduce here a new database called “ImageNet”, a largescale ontology of images built upon the backbone of the WordNet structure. ImageNet aims to populate the majority of the 80,000 synsets of WordNet with an average of 500- 1000 clean and full resolution images. This will result in tens of millions of annotated images organized by the semantic hierarchy of WordNet. This paper offers a detailed analysis of ImageNet in its current state: 12 subtrees with 5247 synsets and 3.2 million images in total. We show that ImageNet is much larger in scale and diversity and much more accurate than the current image datasets. Constructing such a large-scale database is a challenging task.

We describe the data collection scheme with Amazon Mechanical Turk. Lastly, we illustrate the usefulness of ImageNet through three simple applications in object recognition, image classification and automatic object clustering. We hope that the scale, accuracy, diversity and hierarchical structure of ImageNet can offer unparalleled opportunities to researchers in the computer vision community and beyond.

**[10] A. Esteva, B. Kuprel, and S. Thrun**. Deep networks for early stage skin disease and skin cancer classification.

In this study, we investigate what a practically useful approach is in order to achieve robust skin disease diagnosis. A direct approach is to target the ground truth diagnosis labels, while an alternative approach instead focuses on determining skin lesion characteristics that are more visually consistent and discernible. We argue that, for computer-aided skin disease diagnosis, it is both more realistic and more useful that lesion type tags should be considered as the target of an automated diagnosis system such that the system can first achieve a high accuracy in describing skin lesions, and in turn facilitate disease diagnosis using lesion characteristics in conjunction with other evidence. To further meet such an objective, we employ convolutional neural networks (CNNs) for both the diseasetargeted and lesion-targeted classifications.

We have collected a large-scale and diverse dataset of 75,665 skin disease images from six publicly available dermatology atlantes. Then we train and compare both disease-targeted and lesion-targeted classifiers, respectively. For disease-targeted classification, only 27.6% top-1 accuracy and 57.9% top-5 accuracy are achieved with a mean average precision (mAP) of 0.42. In contrast, for lesion-targeted classification, we can achieve a much higher mAP of 0.70.

**[11] S. Ioffe and C. Szegedy**. Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167, 2015.

Training Deep Neural Networks is complicated by the fact that the distribution of each layer's inputs changes during training, as the parameters of the previous layers change. This slows down the training by requiring lower learning rates and careful parameter initialization, and makes it notoriously hard to train models with saturating nonlinearities. We refer to this phenomenon as internal covariate shift, and address the problem by normalizing layer inputs. Our method draws its strength from making normalization a part of the model architecture and performing the normalization for each training mini-batch}. Batch Normalization allows us to use much higher learning rates and be less careful about initialization. It also acts as a regularizer, in some cases eliminating the need for Dropout.

Applied to a state-of-the-art image classification model, Batch Normalization achieves the same accuracy with 14 times fewer training steps, and beats the original model by a significant margin. Using an ensemble of batch-normalized networks, we improve upon the best published result on ImageNet classification: reaching 4.9% top-5 validation error (and 4.8% test error), exceeding the accuracy of human raters.

**[12] G. K. Jana, A. Gupta, A. Das, R. Tripathy, and P. Sahoo**. Herbal treatment to skin diseases: A global approach. Drug Invention Today, 2(8):381–384, August 2010.

Skin diseases are most common form of infections occurring in people of all ages. Skin disorders due to its ugliness and associated hardships are one of the hardest ailments to get accustomed to especially when it is located in a place that is difficult to conceal like the face, even with make up. Most of the skin infections treatment take long time to show their effects. The problem becomes more worrisome if the ailment does not respond to skin disorder treatments. There are not many statistics to prove the exact frequency of skin diseases in this country, but general impression is 10-20 percent of patients seeking medical advice suffer from skin diseases. The skin conditions are prevailant across all parts of the world. Sun is one of the most prominent sources of skin cancer and related traumas.

Herbal supplements are more popular now than ever before. People are looking for new ways to improve their health, and they are turning to natural remedies rather than pharmaceutical drugs more and more. As such, there have been countless studies performed on the therapeutic benefits and applications of various herbs and herbal extracts. Much of this research presents strong evidence that taking herbal supplements in conjunction with a healthy diet and lifestyle can be beneficial. Therefore, many people use natural health supplements to treat various health conditions as well as to promote general well-being. Diseases of the skin account for a great deal of misery, suffering, incapacity and economic loss. Besides this, they are a great handicap in the society, because they are visible.

Fortunately, however, due to recent advances, cutaneous scars can be successfully removed by plastic planning, laser therapy and skin grafting. This review article has highlighted the role and utilities of some medicinal plants on different skin diseases.

**[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Gir- shick, S. Guadarrama, and T. Darrell.**Caffe: Convolu- tional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.

Caffe provides multimedia scientists and practitioners with a clean and modifiable framework for state-of-the-art deep learning algorithms and a collection of reference models. The framework is a BSD-licensed C++ library with Python and MATLAB bindings for training and deploying generalpurpose convolutional neural networks and other deep models efficiently on commodity architectures. Caffe fits industry and internet-scale media needs by CUDA GPU computation, processing over 40 million images a day on a single K40 or Titan GPU (≈ 2.5 ms per image).

By separating model representation from actual implementation, Caffe allows experimentation and seamless switching among platforms for ease of development and deployment from prototyping machines to cloud environments. Caffe is maintained and developed by the Berkeley Vision and Learning Center (BVLC) with the help of an active community of contributors on GitHub. It powers ongoing research projects, large-scale industrial applications, and startup prototypes in vision, speech, and multimedia.

**CHAPTER 3**

**Proposed Model**

Proposed method includes 2 major steps as,

1. Hand gesture Recognition
2. Robot Car Movement based on recognised hand gesture

Dataset Preparation

Background Removal

Training Modified CNN

Prediction Performance

Dataset Preprocessing

Prediction on Webcam data

Robot Car Motion/ Audio Output

Fig.3.1 proposed Method Block Diagram

Proposed method uses deep learning techniques which includes custom CNN (Convolutional Neural Network) for gesture classification and recognition.

Proposed method block diagram is shown in above figure. Proposed steps are discussed in detail below,

* 1. **Creating Dataset**

Dataset created which has 5982 images. There are six classes prepared for controlling the robot car motion. Classes prepared are ‘Backword’, ‘Forward’, ‘Left’, ‘No-Motion’, ‘Right’ and ‘Stop’. Each class has 997 images. The dataset prepared is ‘hand\_dataset’ which has images data which takes 37.5 minutes for training.

Steps followed in Dataset Creation:

1. Start the webcam
2. Generate background Subtractor using KNN
3. Convert Color image in grayscale image
4. Remove Gaussian Blur from image
5. Segment hand by threshold
6. Apply Contour to segmented hand part
7. Create the separate image folder for each hand gesture.

In this dataset creation, finding hand part from contour and saving grayscale image in created folder is important task. For each gesture we have saved 997 images in given folder.

* 1. **Pre-processing dataset**

Input dataset must go through the pre-processing operations to get dataset in standard format. The dataset used in this application is not a normal color image dataset. Color images are converted to grayscale images and resized to given size. Preprocessing helps to get better accuracy as it provides proper format data for classification.

* 1. Background Removal

‘createBackgroundSubtractorMOG2’ is used for creating background subtractor. In this static background is always subtracted from moving background to get only object detection. In this , objects can be detected by leaving background.

* 1. **Training Modified CNN**

Normal image dataset without any background removal takes training time nearly 1 hour whereas the dataset prepared by background subtraction and pre-processing takes only 10 minutes for training to modified CNN and also provides prediction accuracy of 99.6%.

Trained Modified-CNN contains following layers,

1. Convolutional Layer
2. Max-Pooling Layer
3. Dense Layer
4. Flatten Layer
   1. **Prediction using Trained model from webcam Data**

Trained model created used for prediction of new hand gesture from camera. Before prediction the gesture obtained from webcam is converted to grayscale and all the pre-processing techniques, background subtraction techniques are applied to obtained hand gesture image.

Same layers used for CNN training, again used for prediction of hand gesture using CNN.

**3.6 Post-processing**

In post processing, results obtained by hand gesture recognition is transmitted to robot car using arduino. Even obtained results are given to speaker output by generating speech.

**CHAPTER 4**

**Results Analysis**

Proposed modified CNN algorithm and existing random forest classifier algorithms are analysed for gesture recognition to control robot car movement. Dataset with different six types of gestures is prepared which has multiples images inside. Dataset created by us is shown in below figure.

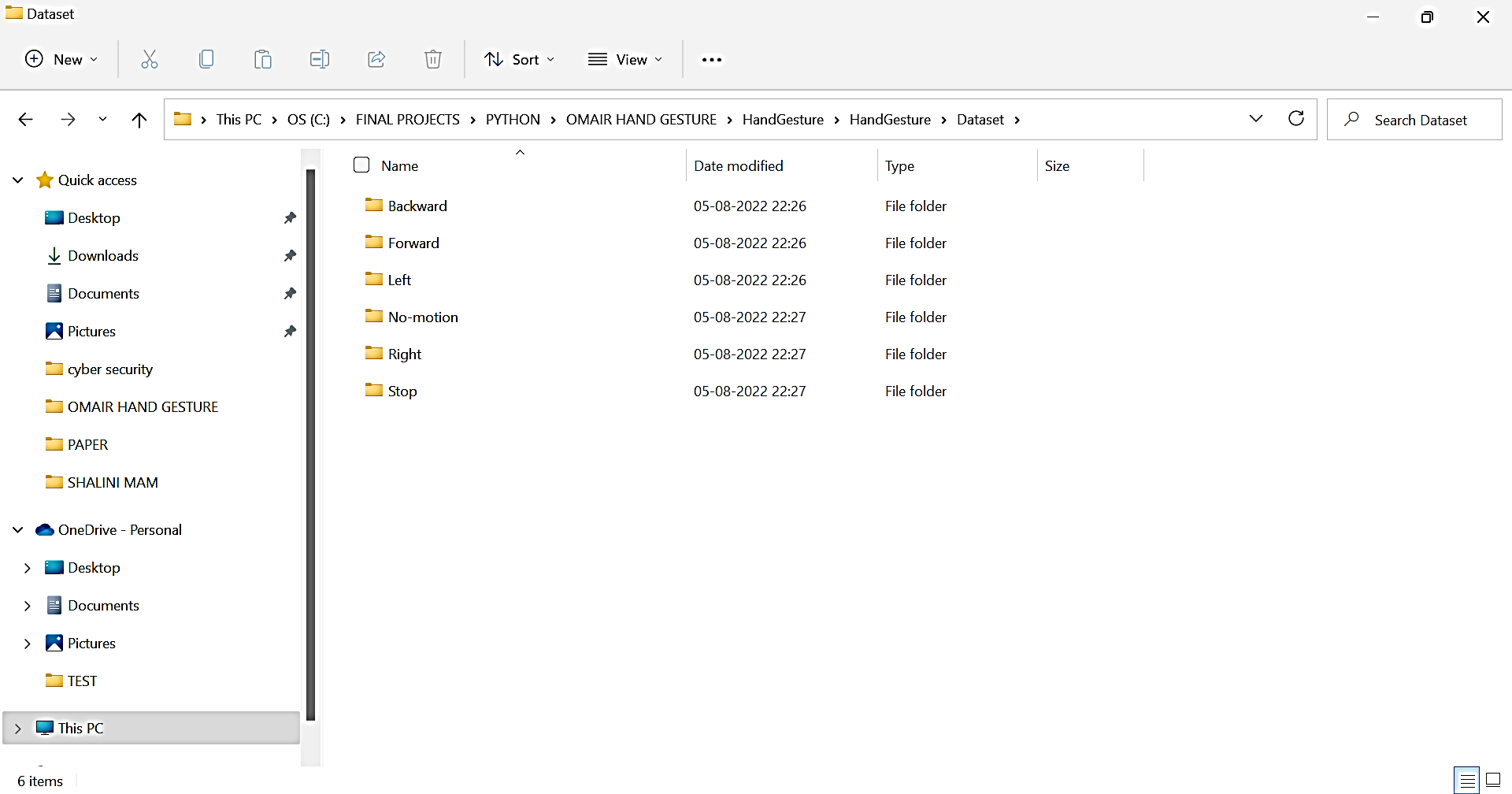


Fig.4.1 Hand Gesture Dataset Created for Training proposed Model

GUI (Graphical User Interface) is prepared in python for easy access to proposed model. Prepared dataset is loaded for training Modified CNN algorithm. Below screen is obtained after trained modified-CNN model.

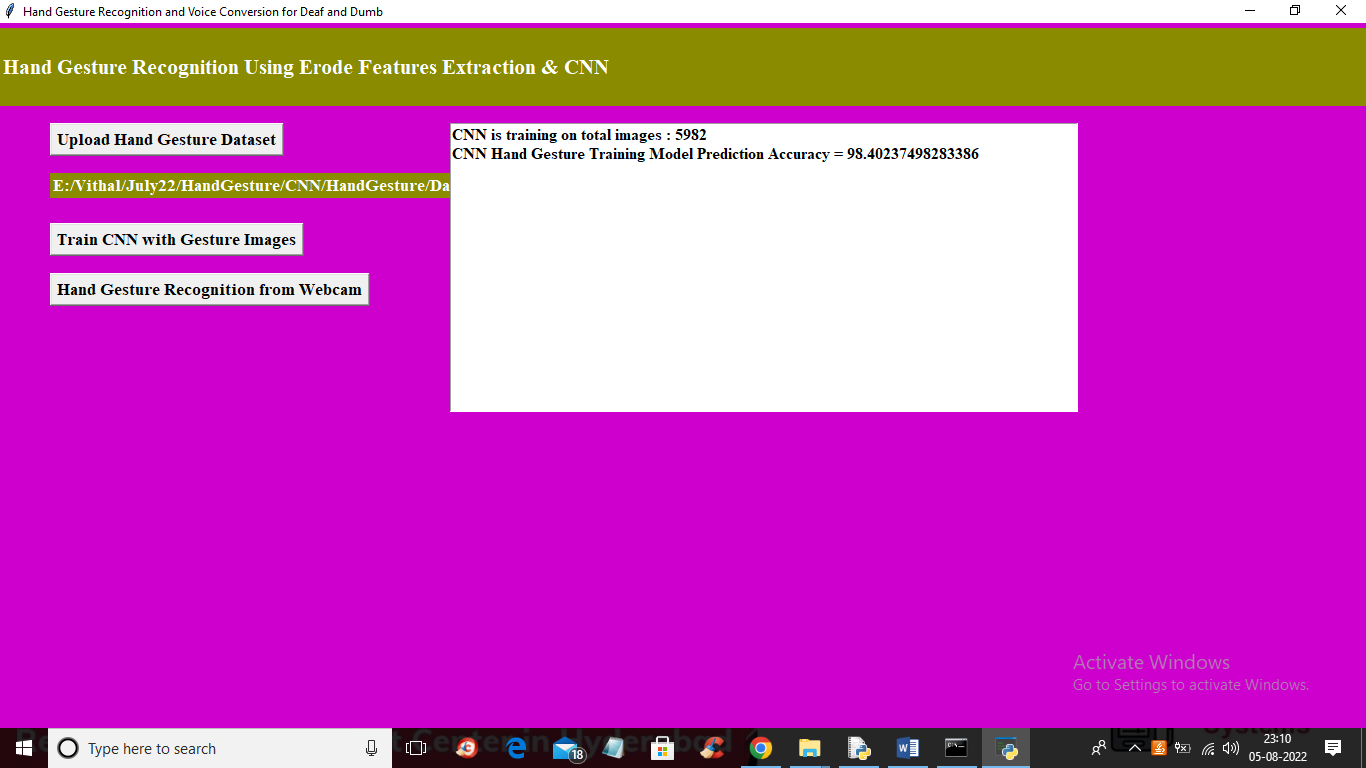


Fig. 4.2 Proposed model performance for prediction

In above results, CNN training model prediction accuracy is 98.4024 %. Original dataset is split to two groups as, 80% of the complete dataset is used for trained modified CNN algorithm while 20% of the complete data is used for prediction of hand gestures using trained model.

In below results, we have took green color bounding box , in which we need to show hand gesture and that gesture will be recognised by proposed modified CNN algorithm.

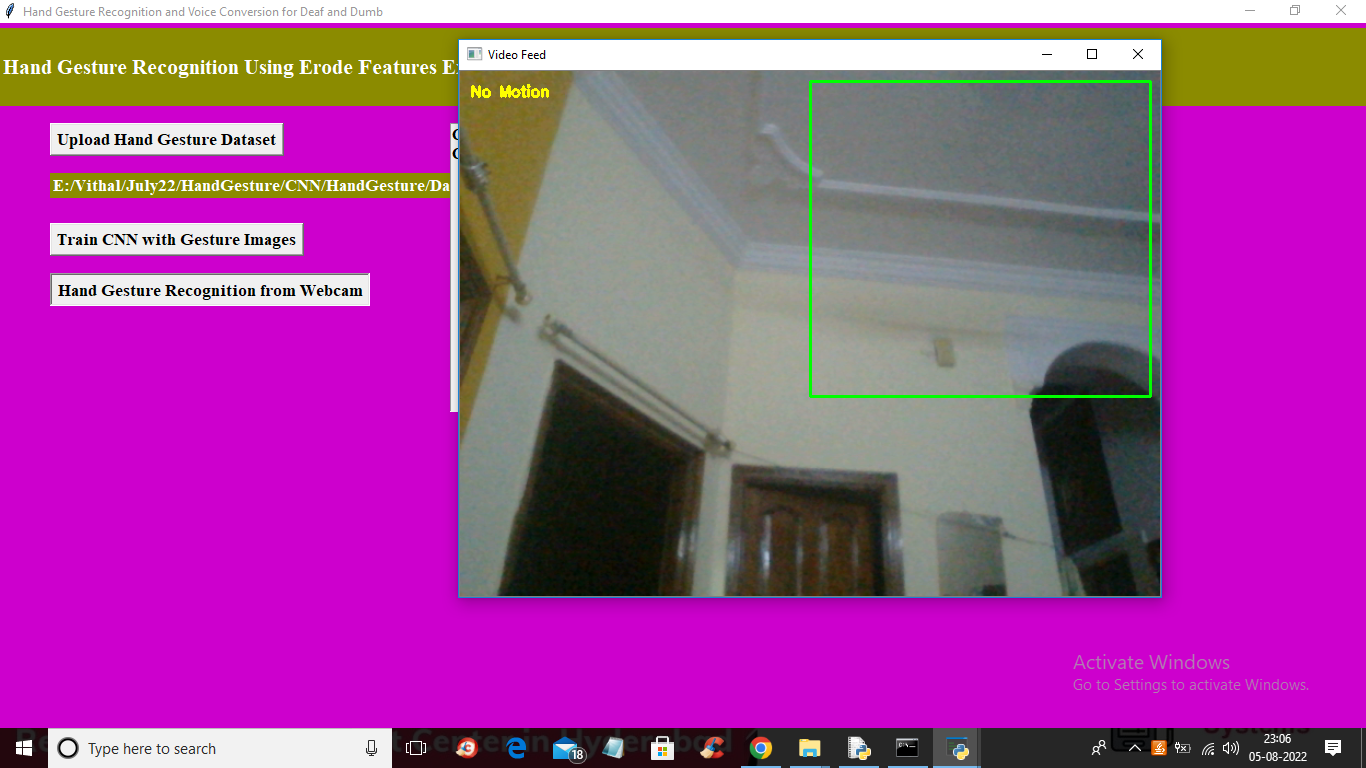


Fig.4.3 webcam data prediction using proposed model

When there is no gesture in front of webcam then recognised hand gesture is ‘No Motion’ based hand gesture. To get this type of prediction, image dataset with stable background, with human face, etc are trained.

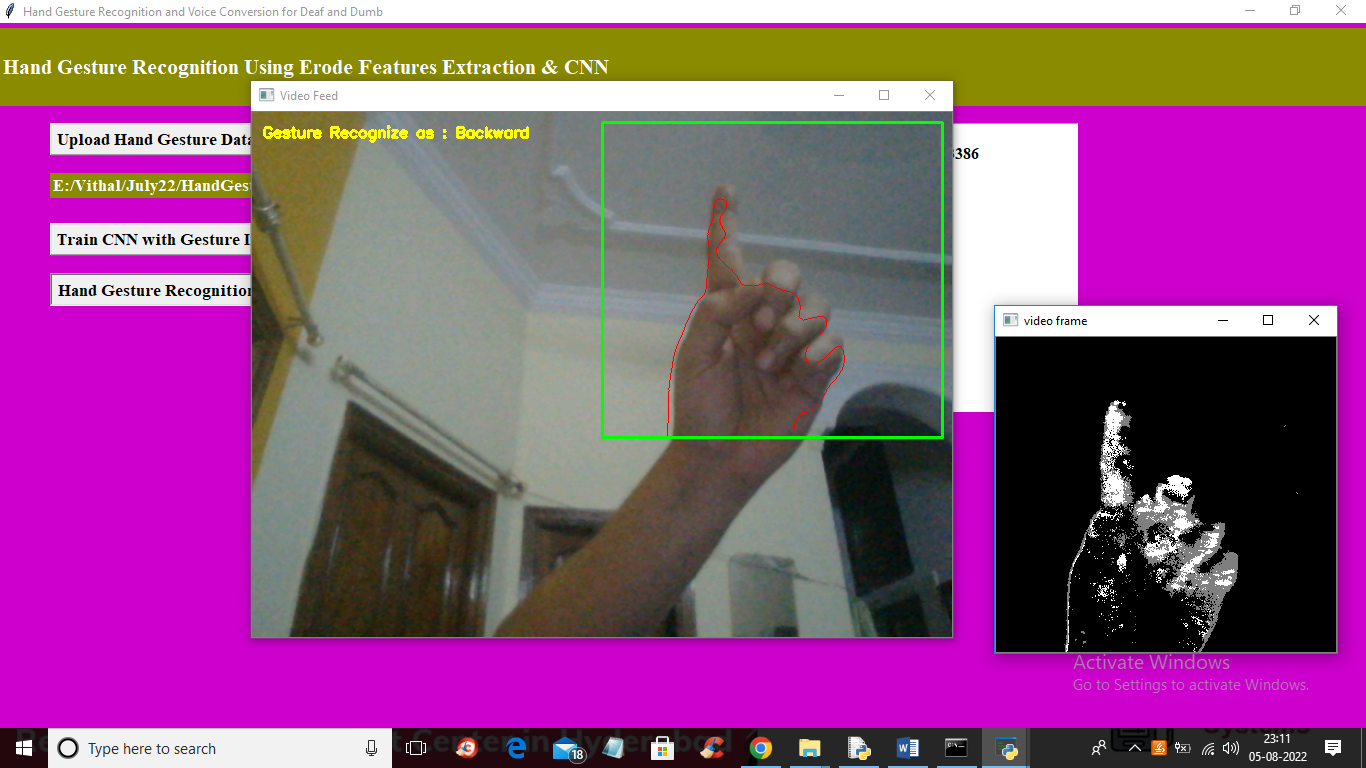


Fig.4.4 webcam data prediction using proposed model

Using proposed model hand gesture recognised is for ‘Backword’ motion. Above figure hand gesture is recognised which send ‘Backword’ motion instruction to robot car. This instruction is transmitted to robot car using Arduino controller.

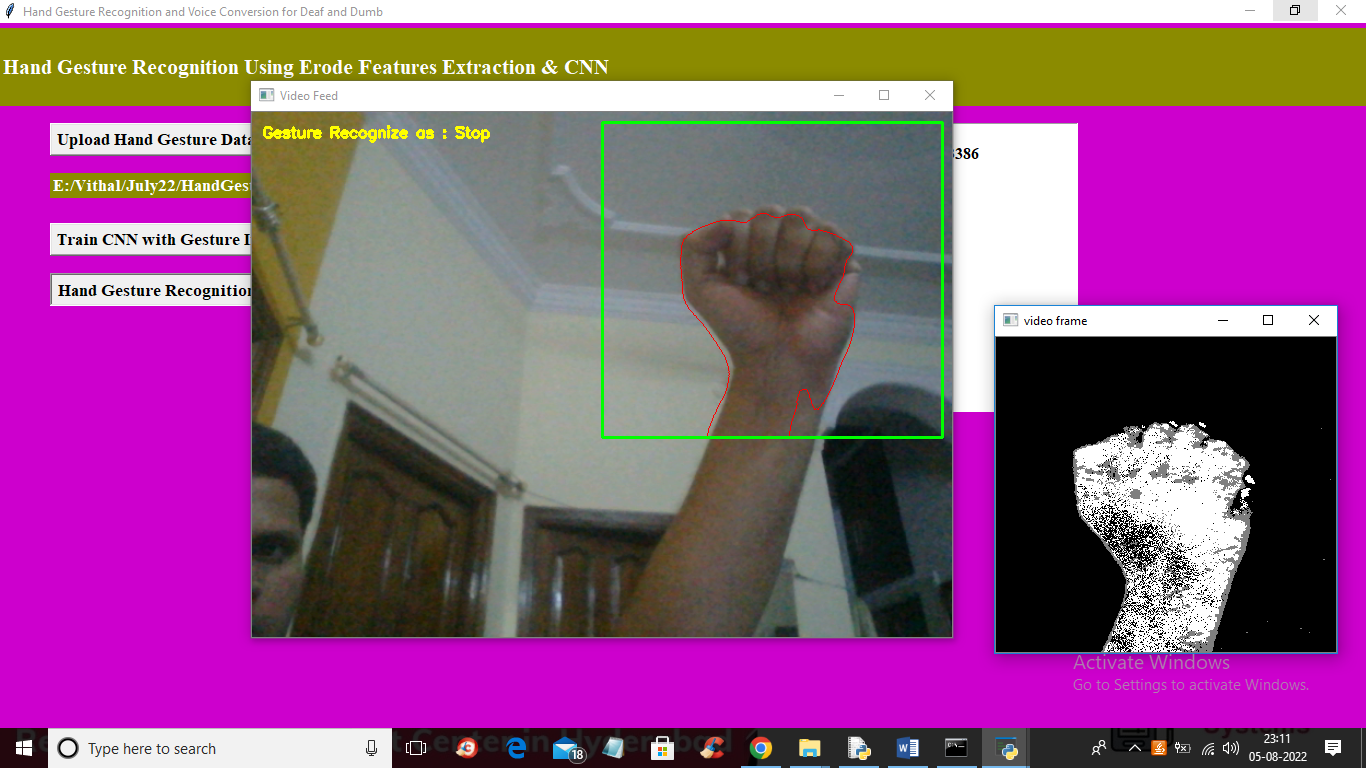


Fig.4.5 webcam data prediction using proposed model

Using proposed model hand gesture recognised is for ‘Stop’ motion. Above figure hand gesture is recognised which send ‘Backword’ motion instruction to robot car. Robot car will be stopped using this hand gesture and the instruction will be transmitted to arduino controller through serial port.

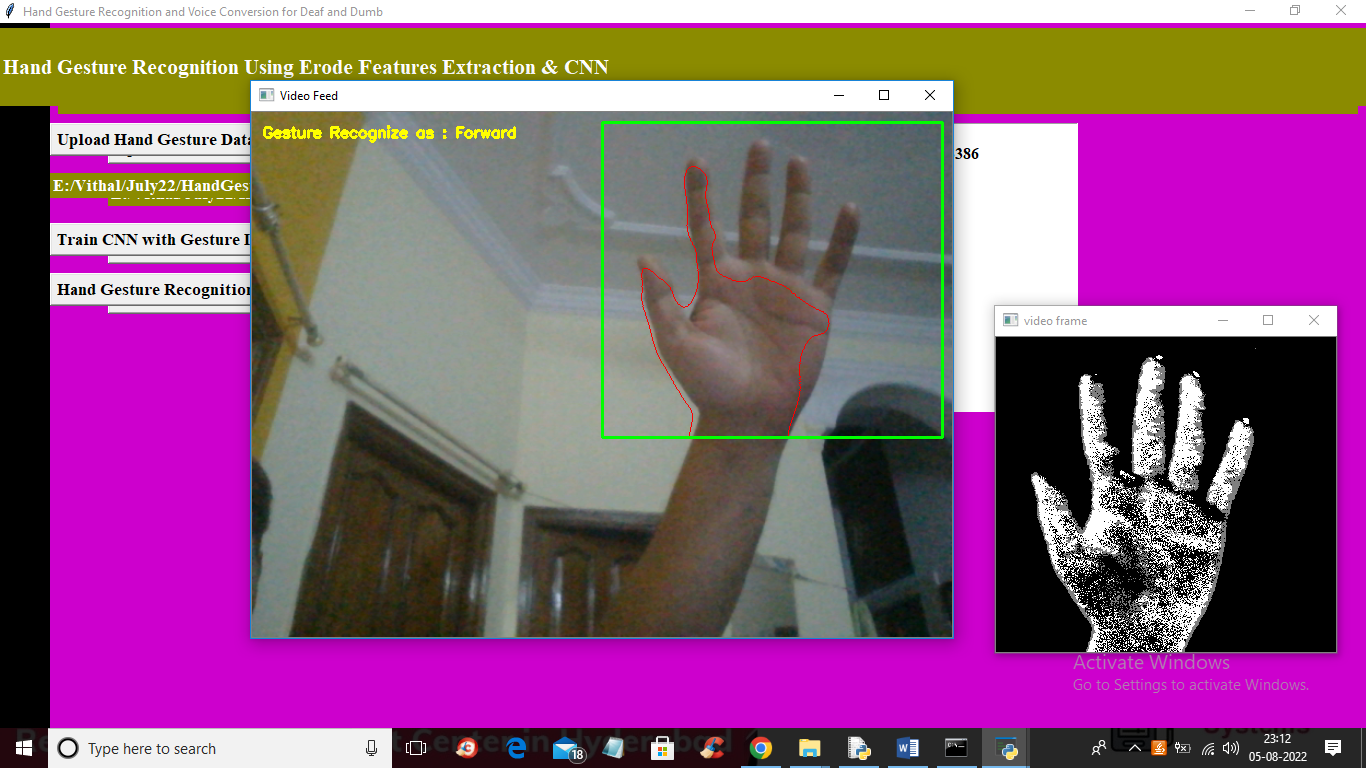


Fig. 4.6 webcam data prediction using proposed model

Using proposed model hand gesture recognised is for ‘Forword’ motion. Above figure hand gesture is recognised which send ‘Forword’ motion instruction to robot car. With this robot car can move straight forward, the instruction is transmitted from laptop/computer to robot car through serial port and Arduino executes the instruction.

**CHAPTER 5**

**Conclusion and Future Scope**

Proposed modified CNN model performs superior compared to state of art technique such as random forest machine learning classifier for controlling movement of robot car. Modified CNN consists of unique structure of layers and these layers modifications is obtained to get higher accuracy. Robot car move in five different directions based on five different gestures and one is ‘No Motion’ gesture. Modified CNN algorithm gives 98.4024 % prediction accuracy while random forest classifier gives only 69% accuracy of real time hand gestures through webcam. In future work, proposed work can be extended to complete real time application using Raspberry pi controller. GPS and some more sensors can be added to get location, environment details. Even obtained results can be published on private cloud.

**REFERENCES**

[1] Saleem Ullah1, Zain Mumtaz1,etAl. (2019) ,An Automated Robot-Car Control System with Hand-Gestures and Mobile Application Using Arduino, © 2019 by the author(s). Distributed under a Creative Commons CC BY license, <https://www.researchgate.net/publication/330182990>

[2] Marc Peral. etAl. (2022) , ‘Efficient Hand Gesture Recognition for Human-Robot Interaction’, Journals & Magazines ,IEEE Robotics and Automation , Volume: 7 Issue: 4.

[3] Tanuja Jha etAl. (2018) , ‘Real Time Hand Gesture Recognition for Robotic Control’, 978-1-5386-5657-0/18/$31.00 c 2018 IEEE

[4] F. Arce, J. M. G. Valdez,” Accelerometer-Based Hand Gesture Recognition Using Artificial Neural Networks” in Soft Computing for Intelligent Control and Mobile Robotics Studies in Computational Intelligence, 2011.

[5] Gesture Controlled Robot using Kinect http://www.eyantra.org/home/projects-wiki/item/180-gesturecontrolled-robot-usingfirebirdv-and-kinect [6] R. Wang, J. Popovic, ”Real-time hand-tracking with a color glove,” ACM Transactions on Graphics,2009.

[7] C.-C. Chang, I.-Y Chen, and Y.-S. Huang, "Hand Pose Recognition Using Curvature Scale Space", IEEE International Conference on Pattern Recognition, 2002.

[8] Chethana N S, Divya Prabha,M Z Kurian, “Static Hand Gesture Recognition System for Device Control”, International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084, Vol 3,April-2015.

[9] Abhijit Nale, Ankush Barkade, Pravin Ekhande, and Shreyas Dayal, “Hand Gesture Recognition for Robotic Car Control”, IJARCST, Vol. 3, Issue 1 (2015).

[10] (DIP) Digital Image Processing by S. Jayaraman, S. Esakkirajan ,T. Veerakumar, August 10, 2013, ISBN: 978-0-07- 014479-8.

[11] Qureshi, M.O.; Syed, R.S. The impact of robotics on employment and motivation of employees in the service sector, with special reference to health care. Saf. Health Work. 2014, 5, 198–202.

[12] Fernando, Y.A.; Mathath; Murshid, M.A. Improving productivity: a review of robotic applications in food industry. Int. J. Rob. Appl. Technol. 2016, 4, 43-62.

[13] Siramshetti, B.K.; Priyanka; Rajitha, C. I won’t fall down; Edge detector robot. Int. J. Sci. Eng. Technol. Res. 2017, 6, 113-116.

[14] Pakdaman, M.; Sanaatiyan, M.M.; Ghahroudi, M.R. A line follower robot from design to implementation: Technical issues and problems. In Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE), Singapore, 26-28 Feb.2010; pp. 5-9.

[15] Singh, R.R.; Das, R.; Pal, K. A novel modus of hand gesture controlled wireless robot. Int. J. Res. Appl. Sci. Eng. Technol. 2017, 5, 641-647.

[16] Fang, B.; Sun, F.; Liu, H.; Liu, C. 3D human gesture capturing and recognition by the IMMU-based data glove. Neurocomputing. 2018, 277, 198–207.

[17] Shi, Y.; Taib, R.; Lichman, S. Gesturecam: A smart camera for gesture recognition and gesture-controlled web navigation. In Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision, Singapore, 5-8 Dec. 2006; pp. 1-6.

[18] Liu, T.; Luo, X.; Liu, J.; Cui, H. Compressive infrared sensing for arm gesture acquisition and recognition. In Proceedings of IEEE International Conference on Information and Automation, Lijiang, China, 8-10 Aug. 2015; pp. 1882-1886.

[19] Dahl, T.S.; Gibbons, P.; Jones, O. Identification and production of simple tactile gestures. In Proceedings of IEEE International Conference on Human-Robot Interaction Workshop on Advances in Tactile Sensing and Touch based Human-Robot Interaction, Boston, Massachussetts, Mar. 2012, doi: 10.13140/2.1.3165.344.

[20] Cao, S.; Yang, P.; Li, X.; Chen, M.; Zhu, P. iPand: Accurate gesture input with smart acoustic sensing on hand. In Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Hong Kong, China, 11-13 June 2018; pp. 1-3, doi: 10.1109/SAHCN.2018.8397157.

[21] Jost, C.; De Loor, P.; Nédélec, L.; Bevacqua, E.; Stanković, I. Real-time gesture recognition based on motion quality analysis. In Proceedings of 7th International Conferences on Intelligent Technologies for Interactive Entertainment (INTETAIN), Turin, Italy, 10-12 June 2015; pp. 47-56.

[22] Ju, M.H.; Kang, H.B. Emotional interaction with a robot using facial expressions, face pose and hand gestures. Int. J. Adv. Rob. Syst. 2012, 9, 1-13.

[23] Chen, C.-H.; Lee, I.-J.; Lin, L.-Y. Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. Res. Dev. Disabil. 2015, 36,396-403.

[24] Vyas, K.K.; Pareek, A.; Vyas, S. Gesture recognition and control. Int. J. Recent Innovation trends Comput. Commun. 2013, 1,575-581.

[25] Vaishnav, P.; Tiwari, S. Accelerometer based hand gesture controlled robot. Int. J. Sci. Res. 2015, 4, 223-226.

[26] Aggarwal, L.; Gaur, V.; Verma, P. Design and implementation of a wireless gesture controlled robotic arm with vision. Int. J. Comput. Appl. 2013, 79, 39-43.

[27] Patel, N.K.; Patel, S.B.; Ammar, M.M. Accelerometer based gesture controlled wheel chair with gps, gsm navigation. Int. J. Innovative Emerging Res. Eng. 2015, 2, 110-113.

**APPENDIX**

**Python**

One of the most popular languages is Python. Guido van Rossum released this language in 1991. Python is available on the Mac, Windows, and Raspberry Pi operating systems. The syntax of Python is simple and identical to that of English. When compared to Python, it was seen that the other language requires a few extra lines. It is an interpreter-based language because code may be run line by line after it has been written. This implies that rapid prototyping is possible across all platforms. Python is a big language with a free, binary-distributed interpreter standard library.

It is inferior to maintenance that is conducted and is straightforward to learn. It is an object-oriented, interpreted programming language. It supports several different programming paradigms in addition to object-oriented programming, including functional and procedural programming. It supports several different programming paradigms in addition to object-oriented programming, including practical and procedural programming. Python is mighty while maintaining a relatively straightforward syntax. Classes, highly dynamic data types, modules, and exceptions are covered. Python can also be utilised by programmes that require programmable interfaces as an external language.

**Python Features:**

**1) Easy:** Because Python is a more accessible and straightforward language, Python programming is easier to learn.

**2) Interpreted language:** Python is an interpreted language; therefore, it can be used to examine the code line by line and provide results.

**3) Open Source:** Python is a free online programming language since it is open-source.

**4) Portable:** Python is portable because the same code may be used on several computer standard **libraries:** Python offers a sizable library that we may utilize to create applications quickly.

**6) GUI:** It stands for GUI (Graphical User Interface)

**7) Dynamical typed:** Python is a dynamically typed language, therefore the type of the value will be determined at runtime.

**Python GUI (Tkinter)**

\* Python provides a wide range of options for GUI development (Graphical User Interfaces).

\* Tkinter, the most widely used GUI technique, is used for all of them.

\* The Tk GUI toolkit offered by Python is used with the conventional Python interface.

\* Tkinter is the easiest and quickest way to write Python GUI programs.

\* Using Tkinter, creating a GUI is simple.

\* A part of Python's built-in library is Tkinter. The GUI programs were created.

\* Python and Tkinter together give a straightforward and quick way. The Tk GUI toolkit's object-oriented user interface is called Tkinter.

\* Making a GUI application is easy using Tkinter. Following are the steps:

1) Install the Tkinter module in place.

2) The GUI application Makes the primary window

3) Include one or more of the widgets mentioned above in the GUI application.

4) Set up the main event loop such that it reacts to each user-initiated event.

Although Tkinter is the only GUI framework included in the Python standard library, Python includes a GUI framework. The default library for Python is called Tkinter. Tk is a scripting language often used in designing, testing, and developing GUIs. Tk is a free, open-source widget toolkit that may be used to build GUI applications in a wide range of computer languages.

**Machine Learning**

Artificial intelligence (AI), which includes machine learning, enables computer systems to learn without being explicitly programmed. It has to do with statistics and applied mathematics. Mike Robert's definition of machine learning. As a computer gathers and learns from the data it provides, it may operate more correctly via machine learning.

For large classes of machine learning, many algorithms are used. We must provide algorithms with more precise data for them to complete certain jobs. In some circumstances, a computer will utilize data to gather information, check its output against the desired outcome, and make necessary corrections.

For instance, when someone texts on a phone, the phone learns about spelling errors and either autocorrects the offending word or suggests a replacement. For many top organizations, machine learning is a critical component of the creation of new products.

ML is an important factor in the operations of many companies, like Facebook and Google. Data science uses machine learning in many different ways. Data scientists rely on ML approaches to carry out their Modeling. Regression and classification are of utmost relevance in data science; hence, the main tool utilized in ML is to accomplish such objectives.

ML applies applicable to practically all phases of data science and is most often associated with the data Modeling phase. Python has been the primary computer language used for data processing. Several Python packages are used in ML settings. The three sections of Python are huge data, optimizing your code, and data files in memory.

**1.6 Types of Machine Learning**

There are three fundamental forms of machine learning: -supervising, semi-supervised, and machine learning

**a) Supervised Machine Learning**

\* That method looks for patterns in the labeled data set to obtain results. Data labelling in supervised learning requires human intervention. To train the algorithm with labeled inputs and the intended output, supervised ML requires human participation. ML under supervision is good for a task like;

**I.** Classify the data using a binary system into two groups.

**II. Multi-classification:** The division of data into more than two categories,

**III.** Modeling imaging continuous value using regression.

**IV. Assembling:** Compiling the estimates from many ML models to provide a precise estimate.

**b) Unsupervised Machine Learning**

\*This method searches for patterns in the data collection without relying on labeled data or human interaction. Data labeling is not necessary for this strategy. ML Unsupervised is effective for tasks like;

**I. Dimensionality reduction:** Reduce the number of variables in the data collection.

**II. Clustering:** Grouping the dataset based on similarities.

**III.** Association mining identifies the item or group of items that commonly appear together in data.

**IV.** Data point identification for anomaly detection in the data set

**c) Semi-supervised Learning**:

\*For this method, you require labeled data. As a consequence, human interaction is also necessary, but the process still moves forward. In this kind of learning, the algorithm is given a tiny quantity of labeled data by data scientists, and as a result, the algorithm gains knowledge about the data set's dimension, which it may then apply to mother del, unlabeled data.

\* There are several contexts in which semi-supervised machine learning (ML) may be used.

**I. Machine translation:** Language conversion using a learning system.

**II. Data labelling:** An algorithm trained on modest amounts of data will automatically apply data labels to enormous collections.

**1.7 Uses of Machine Learning**

\*Machine learning is used in many areas nowadays. The most well-known example is the machine learning recommendation engine that drives a book's news feed. This engine makes an effort to reinforce established patterns in a user's online activity inside a certain Facebook group.

\* The news is appropriately adjusted if a user alters the design and doesn't read anything from that particular group the following week. Applications of machine learning (ML) include business intelligence, human resource information systems, autonomous vehicles, and virtual assistants.

**Advantages:**

• ML helps enterprises in comprehending their clients. ML assists in improving goods in response to client demand by gathering the necessary user data and associating it with shifting behaviour. Some companies' business models are heavily reliant on machine learning, such as Uber, which uses an algorithm to connect drivers and customers. To surface the advertising in searches, Google employs ML.

**Disadvantage:**

• ML might be expensive. High wages for machine learning are a result of data emotions command on the project. These initiatives also often demand expensive software infrastructure.

• In addition to that, when an algorithm is trained on a data set, ML bias might develop. That has flaws in it that might provide erroneous results.

**Steps to choosing the suitable ML model**

The issue is solved by selecting the best ML model, which might take some time. The steps are as follows:

1) For the difficulty with the pure date alignment, the input should be thought about.

2) Gather, label, and prepare the data as appropriate.

3) To put the right algorithms to use and test them to determine how well they perform.

**Libraries Used**

**Pandas:**

\* Pandas is a Python computer language library for data analysis and manipulation. It offers a specific operation and data format for handling time series and numerical tables. It differs significantly from the release3-clause of the BSD license. It is a well-liked open-source of opinion that is utilized in machine learning and data analysis.

**NumPy:**

\* The NumPy Python library for multi-dimensional, big-scale matrices adds a huge number of high-level mathematical functions. It is possible to modify NumPy by utilizing a Python library. Along with line, algebra, and the Fourier transform operations, it also contains several matrices-related functions.

**Matplotlib:**

\* It is a multi-platform, array-based data visualization framework built to interact with the whole SciPy stack. MATLAB is proposed as an open-source alternative. Matplotlib is a Python extension and a cross-platform toolkit for graphical plotting and visualization.

**Scikit-learn:**

\* The most stable and practical machine learning library for Python is scikit-learn. Regression, dimensionality reduction, classification, and clustering are just a few of the helpful tools it provides through the Python interface for statistical modeling and machine learning. It is an essential part of the Python machine learning toolbox used by JP Morgan. It is frequently used in various machine learning applications, including classification and predictive analysis.

**Keras:**

\* Google's Keras is a cutting-edge deep learning API for creating neural networks. It is created in Python and is designed to simplify the development of neural networks. Additionally, it enables the use of various neural networks for computation. Deep learning models are developed and tested using the free and open-source Python software known as Keras.

**h5py:**

\* The h5py Python module offers an interface for the binary HDF5 data format. Thanks to p5py, the top can quickly halt the vast amount of numerical data and alter it using the NumPy library. It employs common syntax for Python, NumPy, and dictionary arrays.