Ministry of Higher Education and
Scientific Research
University of Diyala
College of Science
Department of Computer Science

A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition

A Dissertation
Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By
Hind Ibrahim Mohammed Saba

Supervised By

Assist.Prof. Dr. Jumana W. Salih

2021 A.D. **1443 A.H**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿ قَالُواْ سُبْحَانَكَ لاَ عِلْمَ
لَنَا إِلاَّ مَا عَلَّمْتَنَا إِنَّكَ
أَنتَ الْعَلِيمُ الْحَكِيمُ ﴾

صدق الله العظيم
سورة البقرة (الآية:32)

(Supervisor's Certification)

We certify that this research entitled **"A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition"** was prepared by **(Hind Ibrahim Mohammed)** Under My supervision at the University of Diyala Faculty of Science Department of Computer Science, as a partial fulfillment of the requirement needed to award the degree of Master of Science in Computer Science.

(Supervisor)
Signature:

Name: **Assist.Prof. Dr. Jumana W. Salih**

Date:

Approved by University of Diyala Faculty of Science Department of Computer Science.

Signature:

Name: **Assist.Prof. Dr. Basher Talib Hamid**

Date:

(Head of Computer Science Department)

Linguistic Certification

This is to certify that this thesis entitled **"A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition"** was prepared by **(Hind Ibrahim Mohammed)** at the University of Diyala/Department of Computer Science, is reviewed linguistically. Its language was amended to meet the style of the English language.

Signature:

Name: **Dr.Ghazwan Mohammed Jaafar**

Date: / /

Scientific Amendment

I certify that the thesis entitled **"A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition"** was prepared by **(Hind Ibrahim Mohammed)** has been evaluated scientifically; therefore, it is suitable for debate by the examining committee.

Signature:

Name: Assist.Prof. Dr. Ghadah Kadhim Al-Khafaji

Date: / /

Scientific Amendment

I certify that the thesis entitled **“A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition”** was prepared by **(Hind Ibrahim Mohammed)** has been evaluated scientifically; therefore, it is suitable for debate by the examining committee.

Signature:

Name: Assist.Prof. Dr. Sawsen Abdulhadi Mahmood

Date: / /

Examination Committee Certification

We certify that we have read the thesis entitled **"A Comparison between MSVM and CNN Algorithms in Offline Hand Gesture Recognition"** and an examination committee, examined the student **(Hind Ibrahim Mohammed)** in the thesis content and that in our opinion, it is adequate as fulfill the requirement for the Degree of Master in Computer Science at the Computer Science Department, University of Diyala.

(Chairman)

Signature:
Name**: Prof. Dr. Ziyad Tariq Mustafa**
Date**: / / 2021**

Signature:
Name: **Assist. Prof. Dr. Bashar M. Nema** **(Member)**
Date**: / / 2021**

Signature:
Name: **Dr. Khalid Mohammed Saffer** **(Member)**
Date**: / / 2021**

Signature:
Name: **Asst. Prof. Dr. Jumana W. Salih** **(Supervisor)**
Date**: / / 2021**

Approved by the **Dean** of College of Science, University of Diyala

(The Dean)

Signature:

Name: **Prof. Dr. Tahseen H. Mubarak**

Date**: / / 2021**

Dedication

I would like to dedicate this work to:

To my father may God have mercy on him and mother May God protect her and prolong her life.

To My husband Abbas for his unlimited love, support, endurance, and encouragement.

To my candle, my children Ruqia, and Mohammed.

To my sisters, brothers, and to everyone who helped me from a friend or fellow...

Hind Ibrahim Mohammed

Acknowledgements

My thanks are above all to God Almighty, who guided my steps towards the path of knowledge and without his help and blessing. This thesis would not be advanced or you will see the light.

I express my sincere appreciation to my supervisor's Assist. Prof. Dr. Jumana W. Salih for providing his ideas, inspiration, and continuous support for me during my studies.

I am very grateful to all members and professors of the Computer Science Department at Diyala University.

Finally, I would have never been able to finish my message without especially mention my sisters (Alyaa, Dhamaya, Shaimaa, Esraa, Baidaa, and Nidaa) and support my family and my husband.

Hind Ibrahim Mohammed

ABSTRACT

The process of identifying each letter separately is very important to understand. With this, sign language recognition has become an important technology in artificial intelligence (AI) and machine learning (ML).

This thesis presents two proposed systems for static hand gesture recognition (HGR) based on ML and Deep Learning (DL) algorithms in which several steps are used in the form of phases; image acquisition, image preprocessing, feature extraction, and classification. In the first proposed system, a histogram of oriented gradients (HOG) is utilized for extracting features from each image and then a multi-class support vector machine (MSVM) is applied using the result of the HOG of images to perform the classification process. In the second proposed system, the convolution neural network (CNN) is used through which recognition of static hand gestures is accomplished according to a special structure of this algorithm that consisting of several layers.

The Previous works and researches in that field had a lot of complexity with different accuracy. The obtained results, the second proposed system which adopted DL by using the CNN model outperforms the first system in terms of performance and accuracy, the accuracy rate obtained from the second proposed system was (99.71%) for American Sign Language (ASL) and (99.03%) for Arabic sign language (ArSL), While the accuracy rate obtained from the first proposed system was (95.58%) and (96.16%) for ArSL.

List of Contents

List of Abbreviations

Abbreviations	Description
1V1	One-Versus-One.
1VR	One-Versus-Rest.
AC	Accuracy.
Adam	Adaptive Momentum.
AF	Activation Function.
AI	Artificial Intelligence.
ANN	Artificial Neural Networks.
ArSL	Arabic Sign Language.
ASL	American Sign Language.
BP	Back Propagation.
CM	Confusion Matrix.
CNN	Convolution Neural Network.
CV	Computer Vision.
DA	Discriminant Analysis.
DCC	Deep Convolution Network.
DCT	Discrete Cosine Transform.
DL	Deep Learning.
DNN	Deep Neural Network.
EEG	Electroencephalography.
EOH	Edge Orientation Histogram
FC	Fully Connected.
FN	False Negative.
FP	False Positive.
HCI	Human-Computer Interaction.
HGR	Hand Gestures Recognition.

HMM	Hidden Markov Models.
HOG	Histogram Of Oriented Gradients.
KNN	K-Nearest Neighbor.
LBP	local binary patterns.
LR	Logistic Regression.
ML	Machine Learning.
MLP	Multilayer Perceptron.
MSE	Mean Squared Error .
MSVM	Multi-Class Support Vector Machine.
Nadam	Nesterov-Accelerated Adaptive Moment.
NB	Naïve Bayes.
NN	Neural Network.
PCA	Principal Component Analysis.
PReLU	Parametric Relu.
ReLU	Rectified Linear Unit.
RF	Random Forests.
RGB	Red Green Blue.
Rmsprop	Root Mean Square Propagation.
ROI	Region Of Interest.
SVM	Support Vector Machines.
Tanh	Hyperbolic Tangent Function.
TN	True Negative.
TP	True Positive.

List of Tables

List of Figures

List of Algorithms

Chapter One

GENERAL INTRODUCTION

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Introduction

Gesture recognition in today's technologies is an emerging topic. The major objective of this is to use mathematical algorithms for human-computer interaction (HCI), Typically Gestures may arise from any movement or condition of the body, however generally come from the face or the hand ([1], [2]).

Hand gestures play an important role in communicating human thoughts and feelings, and Sign language is a formal type of hand gestures that are used as a communication device, including visual movements and signs. For the culture of the deaf and speech-impaired. Sign language makes it possible to use several parts of the body such as the fingers, hand, arm, head, torso, or face to communicate details. In the hearing population, sign language is not common, however, although fewer are capable of understanding it. Which creates a real obstacle of contact between both the deaf and the rest of humanity, a question that still has to be completely addressed ([3], [4]).

The deaf and dumb people have been disconnected from the community, and it is impossible for average people to learn sign language. Not only for deaf and dumb individuals, sign language learning has been adopted, but also as a medium for common people to communicate with them ([5], [6]).

Many of the general assets of sign languages around the world are held by ArSL. Its documentation, though, is in a comparatively early process. ArSL also has many nation versions and dialects as most other sign languages [7].

The literature includes several recommended solutions for the recognition of the automated sign language. ArSL, however, has garnered little attention from academics, unlike American Sign Language (ASL) [7].

1.2 Overview of Hand gesture Recognition

Gesture recognition is a sufficient way to understand and follow human language and to assist in communication and interaction between the user and the computer. Gesture recognition is useful for communicating what they cannot communicate with speech or writing text. The best way to express something meaningful is with gestures [1]. Hand gestures recognition (HGR) is still a large research field that is classified according to the meaning of a gesture and the technologies for implementing these gestures [8]. The type of HGR system that is developed is determined by different taxonomies: environmental factors, a system for capturing gestures could be more or less effective, depending on a variety of factors, including the skills of the individual performing the gesture, the effectiveness of the capture systems, the type of gesture (static or dynamic), and the purpose for which the system was designed [9]. Virtual worlds, smart surveillance, sign language translation, medical systems, and other domains all have HGR applications. However, one of the most important applications have developed is sign language based on machine learning (ML) algorithms using hand gestures [10].

Two types of HGR techniques have been described, recognition based on vision and sensor which linking one or more types of sensors, the gesture data is collected using sensor-based recognition. These sensors are connected to a hand that records the hand's positioning and then analyzes the data gathered for gesture recognition. The data glove is an example of sensor-based gesture recognition. Sensors are shown in Figure (1.1), there are certain limitations to base recognition. First, it is necessary to establish the correct hardware, which is really costly. Second, it impedes normal hand gestures. Thus, the weaknesses of sensor-based techniques necessitated the creation of vision-based recognition techniques [11].

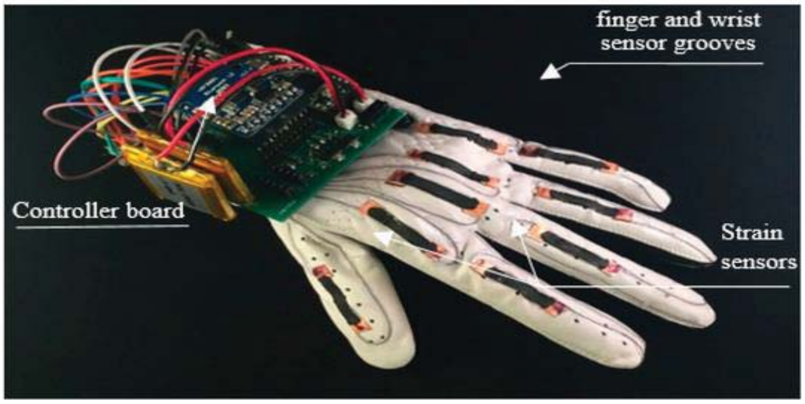

Figure (1.1): Sensor-Based Data Glove [11].

The images of the hands are captured using vision-based methods, which use one or more cameras. Different vision types are shown in Figure (1.2). Stereo cameras, monocular cameras, fish eye cameras, flight time cameras, infrared cameras, and other types of cameras are all available that can be used to capture images. Vision-based techniques use different algorithms for image recognition to achieve hand posture and hand movement. To get the hand position, some vision-based techniques use colored markers. Yet, the limits of vision-based recognition is often influenced by shifts in illumination and cluttered backgrounds, see ([12], [13]).

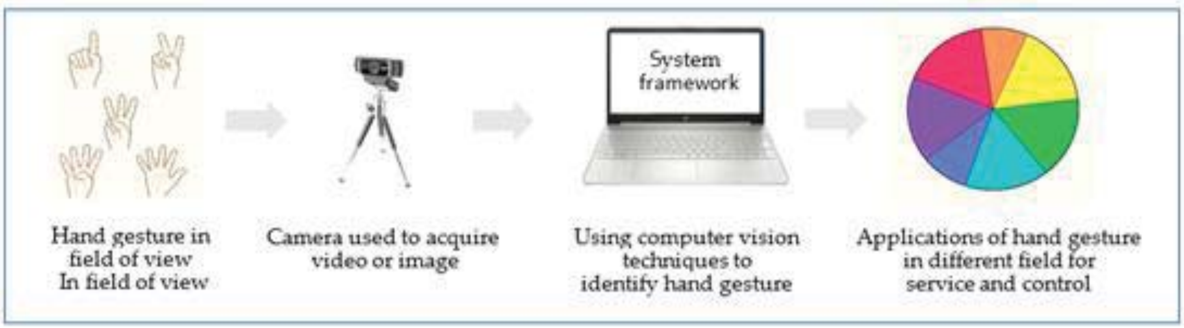

Figure (1.2): Using Computer Vision Techniques [11].

In general, gestures are divided into two categories: static gestures and dynamic gestures. Hand forms are typically used to describe static gestures, while hand motions are used to describe dynamic gestures[14].

HGR systems that rely on geometric features such as fingertips, finger directions, and hand contours are only accurate in situations where they are not occluded or the lighting is not too dark. Shape, color, and texture are present, but they are not sufficiently important for recognition. To define features, images or transformed images are used as the input. The recognizer then implicitly and automatically selects features from the image or transformed image [14].

1.3 Related Work

Advanced tools and strategies have greatly enhanced ML algorithms that is a set of learning methods designed to represent structured data that has successfully been applied to the field of image classification to the extent that they can overwhelm human performance. This section reviews the previous studies that used convolutional neural network (CNN) and Support Vector Machine (SVM) to recognize hand gestures for Different Datasets.

S. Nagarajan and T. S. Subashini (2013) [5]**:** proposed a consistent HGR system for ASL based on the features of Edge Orientation Histogram (EOH) and multi-class support vector machine (MSVM). The database of the image contains a total of 720 images in 24 categories of the American Sign Language (ASL) alphabet, each category contains 30 images. The input sign language alphabets' edge graph count is extracted features and applied to the MSVM for categorization. The average accuracy of the system was 93.75%, the system failed to classify some alphabets. as well as the absence of the letters "J" and "Z" in the data set used because these two gestures are dynamic and include movement.

O. K. Oyedotun and A. Khashman (2016) [15]**:** proposed application of deep learning (DL) for the complete (24 ASL) hand gestures acquired from Thomas Moeslund database to the topic of hand gesture recognition. The complicated role of classifies hand signals at reduced error levels has been demonstrated by more biologically based and deep neural networks, including convolution neural networks (CNN) and stacked denoizing autoencoders.The networks considered are based on the data collected and checked. Recognition scores of 91.33% and 92.83%, the data does not contain the two-letter 'j' and, 'z' and the model can obtain higher accuracy if used in another data set

V. Bheda and N. D. Radpour (2017) [16]**:** proposed a model for using deep convolutional networks (DCN) to classify images, the data set ASL was a collection of 25 images for each alphabet and the digits. CNN architecture consisting of multiple convolutional and dense layers. The accuracy was 67% of letters of the alphabet, and 70% for digits. The number of images for each character can be increased to further improve the system's accuracy.

S. Masood et al. (2018) [17]**:** presented the application of CNN for recognizing hand gestures. used 36 different categories, 26 classes for ASL and 10 classes for Numerals (0-9). Accuracy was 96%. The system failed to classify the zero and 'W’ alphabet as ‘O’ and six respectively.

Reema Alzohairi et al. (2018) [18]**:** aimed to recognize ArSL alphabets automatically using Methodology focused on images Especially, An accurate ArSL alphabet recognizer is being designed through the use of numerous visual descriptors. In the extraction process, the extracted visual descriptors are used as input for the One Versus Rest analysis (1VR). As a result, the ArSL gesture models learned 1VR using histogram of oriented gradients (HOG) descriptors are used. The database contains 30 classes (7 images for each character), image is focused and resize to 200x200 pixels in size. The system accuracy 63.5 %. The system has an issue with a limited dataset for the training algorithm. The number of images for each character can be increased to further improve the system's accuracy.

R. Ahuja et al. (2019) [19]**:** proposed a model that used CNN layers and digital image processing techniques. Open CV was used to track down additional execution methods such as image preprocessing. The archive, which was used to test 24 ASL hand gestures, contains 47,445 photographs, of which 33000 (70%) were used in the training collection and 14445 (30%) were used for testing. The results showed that was accurate at 99.7%. It is attributed that the system detected 24 a letter instead of 26 the absence of the two letters "J" and "Z" in the data set used in the model.

S. Hayani et.al (2019) [20]**:** proposed a model using CNN. This system will detect numbers and letters when fed with a real dataset. Utilized the dataset of images which contained 2,030 images of numbers, and 5,839 images of the 28 different ArSL classes, and the result was an accuracy of 90.02 percent. More accurate results can be obtained by increasing the number of CNN layers and the number of images used for each letter.

T. Goswami and S. R. Javaji (2020) [21]**:** suggested a model that relies on a CNN to recognize and classify hand gestures. The dataset uses 24 classes (27,455 images) to ASL (A-Z), with size (28x28). DL technology based on CNN learns and automatically extracts features to classify each gesture. The proposed model has a test accuracy of 99%. It is attributed that the system detected 24 a letter instead of 26 the absence of the two letters "J" and "Z" in the data set used in the model.

M. M. Kamruzzaman (2020) [22]**:** proposed a system to detect hand signs with CNN automatically to dataset (ArSL), the system was trained for 100 epochs by optimizer with a cost function. The system is then connected to its signature stage, where a hand sign has been translated with 90% accuracy to Arabic speech. That can be improved by increasing the number of images, as only 100 images were used for each letter.

A.Sharma et al. (2020) [23]: proposed a system that used Many various techniques for pretreatments such as HOG, local binary patterns (LBP), and principal component analysis (PCA). This dataset ASL contains 29 classes (3000 image). These methods were successfully implemented to obtain effective results accuracy of Multilayer Perceptron (MLP) 96.96%, K-Nearest Neighbor (KNN) 95.81%, Random Forests (RF) 92.69%, Support Vector Machines (SVM) 85.25%, Logistic Regression (LR) 84.59%, and Naïve Bayes (NB) 72.23%.

Table (1.1): Comparison of Related Works.

No.	Author(s),Year	Ref. No.	Algorithm for Classification	Dataset Size (Images Number)	Accuracy
1.	S. Nagarajan and T. S. Subashini (2013)	[5]	MSVM	720 images in 24 categories ASL	93.75%,
2.	O. K. Oyedotun and A. Khashman (2016)	[15]	CNN	1440 for training , 600 for testing ASL	92.83%
3.	V. Bheda and N. D. Radpour (2017)	[16]	CNN	650 Images , 25 images from 5 people for each alphabet ASL	67%
4.	S. Masood et al. (2018)	[17]	CNN	2524 ASL gestures	96%
5.	Reema Alzohairi et al. (2018)	[18]	MSVM	210 ArSL	63.5 %.
6.	R. Ahuja et al. (2019)	[19]	CNN	47,445 images for 24 classes ASL	99.7%
7.	S. Hayani et.al (2019)	[20]	CNN	5839 images of 28 class ArSL	90.02%
8.	T.Goswami and S. R. Javaji (2020)	[21]	CNN	27,455 images for 24 classes ASL	99%
9.	M. M. Kamruzzaman (2020)	[22]	CNN	100 images for each alphabet (32 classes) ArSL	90%
10.	A.Sharma et al. (2020)	[[23]	MSVM	3000 images ASL	85.25

1.4 Problem Statement

About 70 million people (deaf and dumb) use sign language as their first or mother tongue all over the world and unfortunately they cannot communicate with the general public because they do not understand the meaning of sign language gestures and on the other hand they are unable to understand natural language.

Indeed, it is very important to support this category of society in view of the great development in the world of technology and software, and as a result, it is necessary to prepare systems capable of translating signals into text or speech. If these systems are put in place, it will greatly help them to understand what is going on around them in an easy and simple way.

For the sake of the above, several researchers have proposed the development and implementation of automated systems or human-computer interaction (HCI) to help deaf people and the general public communicate.

1.5 Aim of the Thesis

This thesis aims to build a strong to recognize hand gesture system for ASL and ArSL sign language to help the deaf and dumb people more easily with computer vision applications using multi-class support vector machine (MSVM) classing was designed and applied as the most important algorithm of ML algorithms in the first proposed system. Furthermore, the CNN model is utilized in the second proposed system which is the most powerful algorithm for DL for making a comparison between these techniques to determine the best one in achieving a high degree of accuracy.

1.6 The Organization of the Thesis

This thesis is organized into four chapters, in addition to the one already described, and is structured as follows:

Chapter Two describes of the theoretical background of the main systemses that used for the hand gesture recognition based on ML.

Chapter Three presents the details of the proposed recognition and classification algorithms that are used to design the proposed system and the implementation of each one.

Chapter Four gives the experimental results obtained from the implementation of the proposed system.

Chapter Five discusses results, conclusions and lists a number of suggestions for future studies.