



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: <https://doi.org/10.22214/ijraset.2023.57544>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Recognition of Sign Language by Using the Implementation of Convolutional Neural Networks

Boddu Vikram¹, Dr. M Sumender Roy²

¹M.Tech Department of Computer Science and Engineering, Lenora College of Engineering, Rampachodavaram, Andhra Pradesh, India

²Professor & Head of Department of Computer Science and Engineering, Lenora College of Engineering, Rampachodavaram, Andhra Pradesh, India

Abstract: *The discipline of Sign Language Recognition (SLR) is primarily concerned with the development of techniques that may translate sign language into written or spoken language. The primary purpose of this field is to make it easier for people who are deaf-mute to communicate with the general public. Despite the inherent difficulties that are brought about by the activity's intricacy and the enormous repertory of manual movements that it demands, this specific activity has a significant influence on the larger social setting. The approaches that are now in use for sign language recognition (SLR) are dependent on features that are developed manually in order to portray the motion of sign language. These features are then used in the process of developing classification models. On the other hand, the issue of building dependable features that are capable of properly responding to the wide variety of hand gestures is a considerable difficulty. In this investigation, we suggest a unique convolutional neural network (CNN) architecture as a possible solution to the issue that was described before. One of the capabilities of the Convolutional Neural Network (CNN) is its capacity to independently extract discrete spatial-temporal properties from unprocessed video streams. This eliminates the need for previous information and eliminates the demand for feature building. CNN utilizes many video feeds that include color information, depth cues, and bodily joint locations. These streams are included into the network. A successful integration of color, depth, and trajectory information is achieved by the use of these streams as input. The effectiveness of the Convolutional Neural Network (CNN) is going to be improved as a result of this activity. Through the application of the proposed model to a real dataset received from Microsoft Kinect, the validity of the model is shown. The results of our investigation indicate that its performance is better to that of conventional techniques that are dependent on attributes that are manually developed.*

Keywords: *Sign Language, CNN, Color Integration, Sign Language Recognition*

I. INTRODUCTION

Those who are deaf or have hearing issues frequently use sign language as a form of communication. It includes the utilization of a variety of hand forms, body motions, and even facial expressions in order to transmit information. As a result of the intrinsic difficulties of efficiently integrating information from hand shapes and body movement trajectories, the job of detecting sign language continues to be a significant challenge. A competent recognition model is presented in this work with the intention of easing the translation of sign language into written or vocal representations to be used in communication. Through the use of sign language, the purpose of this idea is to make it easier for those who have hearing impairments to communicate with others who do not have such impairments.

The creation of descriptors that are capable of accurately describing hand forms and motion trajectories is the one aspect of sign language recognition that presents the greatest challenge from a technical point of view. There are numerous essential processes involved in the process of describing the form of the hand. To begin, it requires keeping an eye on the portions of a video stream that include hand areas. In order to do this, it is necessary to recognize and trace down the areas of the movie that include hands. Second, it necessitates the separation of hand-shaped pictures from the several complex backgrounds that are present in each frame. To complete this stage, you will need to isolate the hand form from the components that are around it in the picture. Last but not least, it entails overcoming issues that are associated with gesture recognition. The resolution of any issues that may emerge throughout the process of recognizing and interpreting hand signals is included in this. A connection may also be formed between the process of monitoring crucial points and curve matching and the trajectory of motion. This correlation may be developed further. The production of a solid result for single-label recognition (SLR) continues to be difficult, despite the fact that substantial research efforts have been made on the issues of hand and body joint variation and occlusion.

To add insult to injury, the issue of merging hand-shape attributes with trajectory information is a tough and demanding problem to handle. Convolutional neural networks (CNNs) have been developed by researchers in order to overcome these concerns. CNNs have the power to include hand gestures, motion trajectories, and facial expressions shown by the participants in a way that is more intuitive. During the same time period, we include color pictures, depth images, and body skeleton images as input. Microsoft Kinect is the source of every single image. In contrast to the traditional method of using color photos as input to networks, which is discussed in references [1, 2], this technique utilizes colored pictures.

II. TECHNOLOGIES USED

Ille Python is largely considered to be the most popular high-level programming language at the moment, and it is used often in a variety of different application areas.

Python is a flexible programming language that supports many different programming paradigms, including procedural and object-oriented programming. When compared to applications built in other programming languages, such as Java, Python scripts are often more succinct for the most part.

Programmers are required to write less code since the indentation requirement of the programming language requires them to do so. This helps to ensure that their work is readable.

Numerous significant technical companies all over the world, including but not limited to Google, Amazon, Facebook, Instagram, Dropbox, and Uber, use the programming language Python.

The extensive collection of standard libraries that Python offers is one of the programming language's most prominent points of strength. Users are able to do a broad variety of jobs thanks to the extensive functionality that these libraries provide, which includes, but is not limited.

A. Extensive Libraries

Python is a well-known programming language that can be learned in conjunction with a complete library. The library under consideration has a broad diversity of code that may be used for a variety of applications. These applications include, but are not limited to, regular expressions, web browsers, documentation generation, unit testing, threading, databases, CGI, email, image processing, and a number of other applications. Taking this into consideration, it seems that the whole human authorship of the code for the operation is not required.

B. Extensible

According to what was shown in the part before this one, Python has the capability of being extended to include additional programming languages. The possibility of including certain sections of one's code by using programming languages such as C++ or C is something that might be taken into consideration. The use of this specific resource shows itself to be quite beneficial, especially in the context of project management.

C. Embeddable

In Python, the embeddability feature is one of the factors that leads to the language's greater extensibility. The embedding of Python code into the source code of another programming language, such as C++, is a topic that should be investigated further. As a result of this development, the existing code, which was initially developed in a different programming language, may now be improved by the addition of scripting capabilities. This is a consequence of the progress.

D. Improved Productivity

The simplicity of the programming language, in addition to its extensive library support, may be a contributing factor to the enhanced productivity of programmers. This is especially true when contrasted with languages such as Java and C++. It is advised that the amount of written output be reduced and that the completion of the job be prioritized.

E. IOT Opportunities

Python's growing popularity as a fundamental programming language for developing platforms like the Raspberry Pi is a strong indicator that the Internet of Things (IoT) is likely to see a more positive trajectory in the near future. One possible way to establish a connection with the material world is through the use of this tactic in language.

F. Simple and Easy

Within the realm of Java programming, it may be of the utmost importance to create a class that performs the function of supporting the production of the textual phrase "Hello World." The inclusion of a print statement is the sole component that is required for Python to function properly. The procedure of gathering knowledge, gaining an understanding of it, and putting it to use via the use of code seems to be a rather straightforward approach. Due to the fact that this particular component is present, it is possible that some people will have issues transitioning to languages that are more verbose, such as Java. There is a possibility that learning Python programming abilities will provide a significant amount of challenge.

G. Readable

Due to the fact that Python is recognized for having a concise syntax, the process of decoding Python code is comparable to the act of reading English. There is a possibility that this quality is associated with the ease of learning, comprehending, and coding. Furthermore, the use of curly brackets for the building of blocks is considered superfluous by virtue of the fact that necessary indentation is needed on every single occasion. The implementation of this modification leads to an increase in the degree of readability that is present in the piece of code.

H. Object-Oriented

The programming language that is currently under study is capable of supporting both procedural and object-oriented programming paradigms. When compared to classes and objects, functions have the advantage of facilitating the reuse of code, but classes and objects give us the capacity to accurately simulate events that occur in the real world. Using a class makes it possible to include data and functions in a single object. The use of a class makes this possible.

I. Free and Open-Source

Python is a programming language that is freely available to everyone, as was said before. In spite of this, it is essential to recognize that Python may be acquired without the need to commit any financial resources. Additionally, it is feasible to get access to the source code of Python, edit it, and distribute it to other people. This is available to everyone who is interested. There are a wide variety of libraries that may be browsed and downloaded in order to improve the efficiency with which one completes their job.

Portable

The situation in which a person writes code for a project using a programming language such as C++ is one in which it is possible that certain adjustments may need to be made in order to simplify the execution of the code on a different platform. Unlike other programming languages, Python does not fall into the same group. It is only necessary to write the code once, and it may be executed in any place that the user chooses. This strategy is sometimes referred to by the acronym WORA, which stands for "Write Once, Run Anywhere." Nevertheless, it is of the utmost importance to exercise sufficient vigilance in order to avoid the introduction of any functionality that is reliant on the system.

III. PROPOSED METHODOLOGY

The full name of Unified Modeling Language, which serves as an acronym to distinguish it from other modeling languages, is UML. The Unified Modeling Language, often known as UML, is a contemporary method that is used for the goal of modeling and making descriptions of software systems. When modeling business processes, this technique is often used because of its widespread application. The diagrammatic representations of software components serve as the essential framework upon which the system is constructed as a whole. In line with a time-honored proverb, it is hypothesized that the visual depiction of an idea or concept has a communication capacity that is comparable to that of a large quantity of textual explanation, which is believed to be on the order of one thousand words. Visual representations have the potential to improve our understanding of any faults or issues that may exist in software or business procedures. The Unified Modeling Language (UML) came into being as a consequence of the major difficulties that were encountered throughout the process of developing and documenting software. A great number of different approaches for the modeling and documentation of software systems came into existence throughout the 1990s. The realization that there was a need for a standardized method of graphically expressing complex systems was the impetus for the development of the Universal Modeling Language, often known as UML. Three software developers working at Rational Software worked together on the design of this product during the years 1994 and 1996. This development took place while they were employed there. Subsequently, in the year 1997, a decision was made to establish it as the prevalent standard, which has continued to this day, although with a few alterations that are not very significant.

The primary goals in the design of the UML are as follows:

- 1) It is important that users have access to a visual modeling language that is not just capable of engaging in expressive modeling but also easily accessible for use. Because of this, they will be able to construct and distribute models that have a large amount of value.
- 2) The availability of tools that promote extension and specialization should be made available in order to enable the growth of essential notions.
- 3) Upholding one's independence from certain programming languages and development methodologies is of the utmost significance.
- 4) In order to make the modeling language easier to grasp, you need to provide a formal framework.
- 5) Encourage the expansion and development of the product market for object-oriented tools.
- 6) Provide assistance in the deployment of advanced development ideas over the whole of the development lifecycle. These concepts include collaborations, frameworks, patterns, and components.

A. AutoencoderArchitecture

An autoencoder can essentially be divided up into three different components:

The encoder, a bottleneck, and the decoder.

The encoder portion of the autoencoder is typically a feedforward, densely connected network. The purpose of the encoding layers is to take the input data and compress it into a latent space representation, generating a new representation of the data that has reduced dimensionality. The code layers, or the bottleneck, deal with the compressed representation of the data. The bottleneck code is carefully designed to determine the most relevant portions of the observed data, or to put that another way the features of the data that are most important for data reconstruction. The goal here is to determine which aspects of the data need to be preserved and which can be discarded. The bottleneck code needs to balance two different considerations: representation size (how compact the representation is) and variable/feature. relevance. The bottleneck performs element-wise activation on the weights and biases of the network. The bottleneck layer is also sometimes called a latent representation or latent variables. The decoder layer is what is responsible for taking the compressed data and converting it back into a representation with the same dimensions as the original, unaltered data. The conversion is done with the latent space representation that was created by the encoder. The most basic architecture of an autoencoder is a feed-forward architecture, with a structure much like a single layer perceptron used in multilayer perceptron's. Much like regular feed-forward neural networks, the auto-encoder is trained using back propagation.

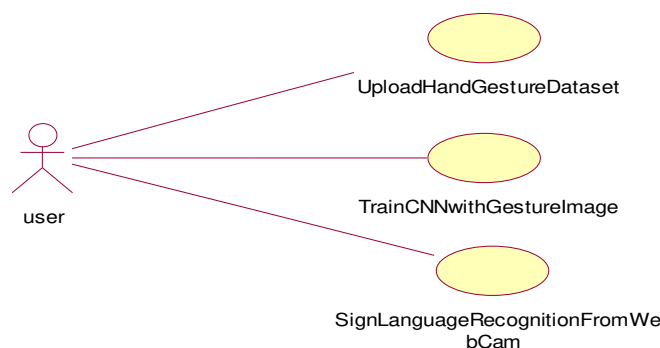


Fig 1. Case Diagram

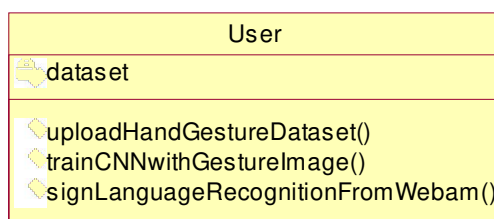


Fig 2. Class

B. Digital Processing

Picture processing is a technique that involves a series of operations that are carried out on a picture with the intention of either improving the image's visual quality or obtaining useful information from the image.

In its most basic form, image processing is characterized by the examination and modification of a digital photograph with the intention of improving the image's overall quality. The core of image processing, which will be the subject of our conversation, is encapsulated in this explanation.

On the subject of digital technology, the following:

While the variables x and y are used to describe spatial (plane) coordinates, the amplitude of the coordinate pair (x, y) is what determines the intensity or gray level of the picture at a particular location. It is possible to describe a picture as a mathematical function $f(x, y)$ that acts in two dimensions, with x and y denoting spatial coordinates. This definition is applicable to most images. To put it another way, a picture may be thought of as a matrix that comprises two dimensions, or in the case of colorful pictures, it can be thought of as a matrix that contains three dimensions. $f(x, y)$ is a mathematical function that delivers the pixel value at each specified place inside an image. This function is responsible for determining the matrix. The pixel value may be used to provide the brightness of a particular pixel, in addition to the chromaticity that corresponds to that pixel.

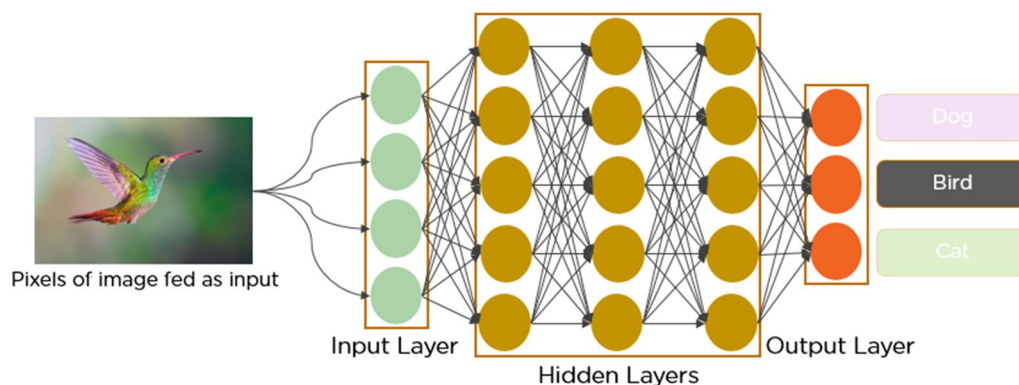
A picture is used as the input for the field of image processing, which may be thought of as a subfield of signal processing. The end product of image processing might be a picture in and of itself, or it can be a collection of qualities that are chosen carefully according to the requirements that are linked to the image.

In the field of image processing, the three stages that are generally recognized are universally accepted.

- Firstly, the image is being introduced.
- The application of analysis and processing tools to the picture
- The output may include several forms, such as a changed photograph or a report derived from image analysis, both serving as illustrative examples.

C. How does a computer read an image

The capacity of deep learning to properly manage enormous amounts of data is primarily responsible for its effectiveness as a powerful tool, which has been shown over the course of the last several decades. In the field of pattern recognition, there has been a discernible change in emphasis toward the use of hidden layers, which has surpassed the attention that was previously devoted to traditional approaches. Deep neural networks that are widely used within the field of deep learning are known as convolutional neural networks (CNNs), which are also often referred to as CNNs or ConvNets. This assumption is particularly valid when considered in the context of applications that include computer vision techniques.



Since the 1950s, researchers have been actively involved in the quest to construct a system that is capable of comprehending visual information. This pursuit has been especially focused on the area of artificial intelligence, which was still in its early stages at the time. Following that, professionals in this area started referring to this specific expertise as "computer vision"; this was a very recent development. A significant achievement was accomplished in the field of computer vision in 2012, when a group of researchers from the University of Toronto developed an artificial intelligence model that outperformed the most advanced algorithms for image recognition by a significant margin. This milestone was a significant achievement in the field of computer vision.

The computer vision competition that ImageNet hosted in 2012 saw success for the artificial intelligence system known as AlexNet, which bears Alex Krizhevsky's name as its primary designer. An outstanding 85 percent accuracy rate was achieved in this particular endeavor. Seventy-four percent was the score that the runner-up received on the examination.

One of the most important components of AlexNet was a particular category of neural networks known as convolutional neural networks (CNNs). These CNNs were developed with the intention of imitating human vision in a general sense. As a result of the fact that convolutional neural networks (CNNs) have become an essential component in a variety of computer vision applications, they are now an essential component in online computer vision courses. Within the context of deep learning, let us now investigate the operational procedures of convolutional neural networks, also known as CNNs.

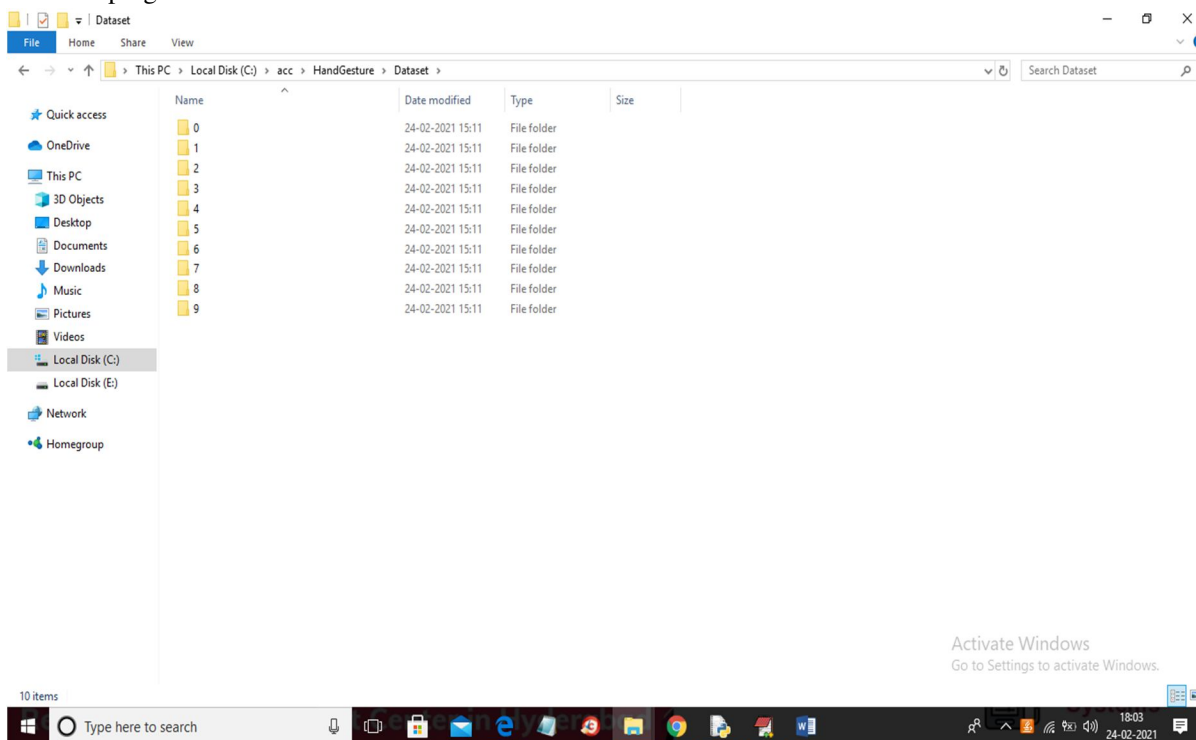
IV. EXPERIMENTAL ANALYSIS

The results of a number of studies have shown that convolutional neural networks (CNNs) that have been trained on well-known datasets like ImageNet have limitations in terms of object identification when they are presented with a variety of lighting conditions and different angles of view.

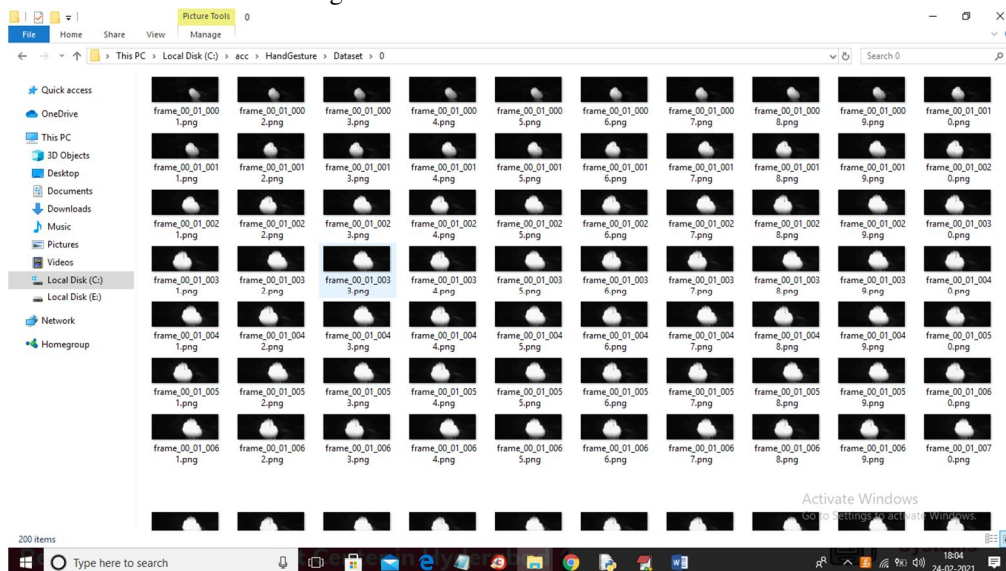
On the basis of the information presented above, what are some of the conclusions that may be drawn about the use of convolutional neural networks (CNNs)? In spite of this, it is indisputable that convolutional neural networks have been the driving force behind a paradigm shift in the area of artificial intelligence. It is important to remember, however, that these networks do have certain limitations that they must adhere to. Convolutional neural networks, often known as CNNs, are widely used in a wide variety of computer vision fields, such as face recognition, image search and manipulation, augmented reality, and a number of other applications. As a demonstration of the exceptional and advantageous progress that has been made in this area, the breakthroughs that have been made in convolutional neural networks (ConvNets) serve as a testimonial. On the other hand, it is essential to recognize that we have not yet arrived at a stage where we are able to effectively imitate the fundamental components of human intellect.

A. Screen Shots

We are using CNN to identify the movement of hand gestures among the participants in this research project so that we may accomplish the goals that have been set for this research endeavor. The photos that are being used to educate CNN may be seen in the screen shots that are provided below. CNN is now getting training from these photographs. These photos are being used in CNN's educational program.



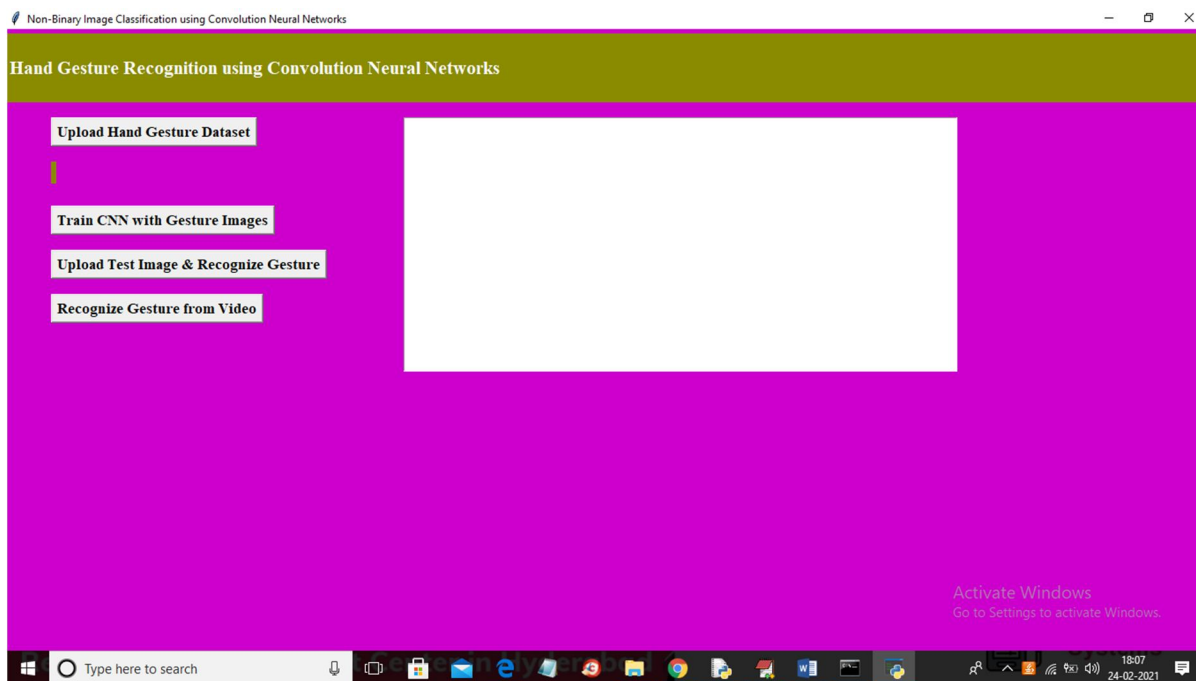
There are ten distinct sorts of hand gesture pictures that are shown in the image that is being displayed. Proceed to any directory on the computer in order to have access to these images.



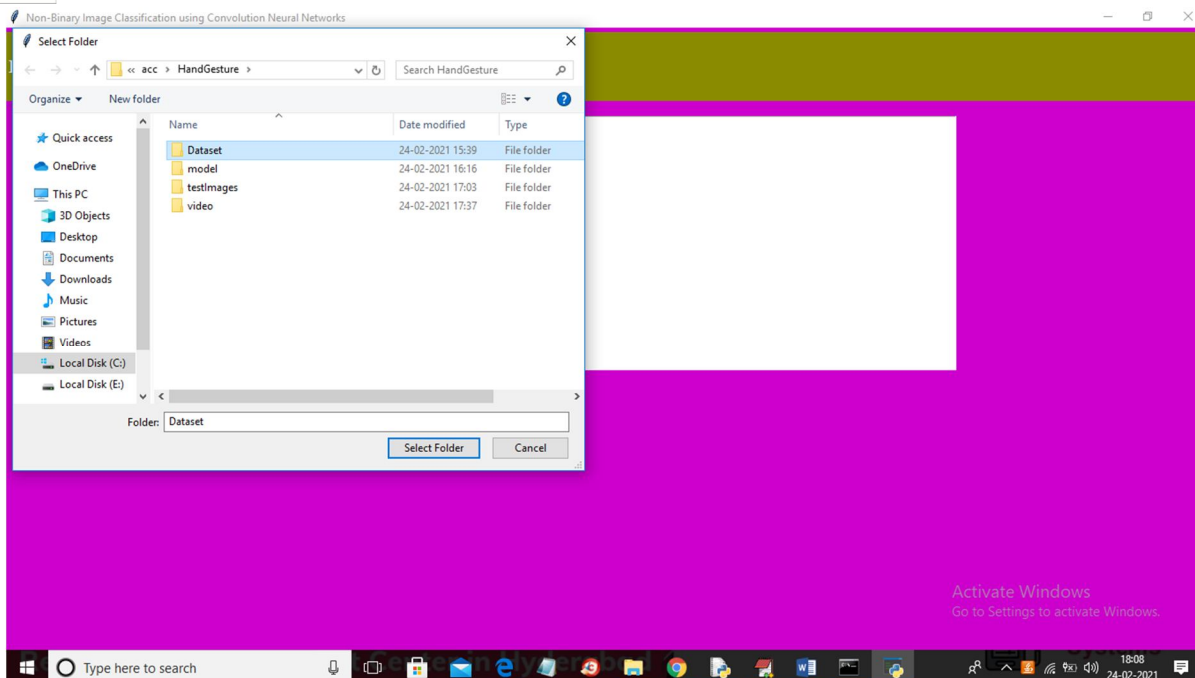
Images that were taken from the folder that was designated "0" are shown in the visual representation that can be seen above. Additionally, it is possible to see a wide range of images that have been taken from various directories.

B. Screen Shots

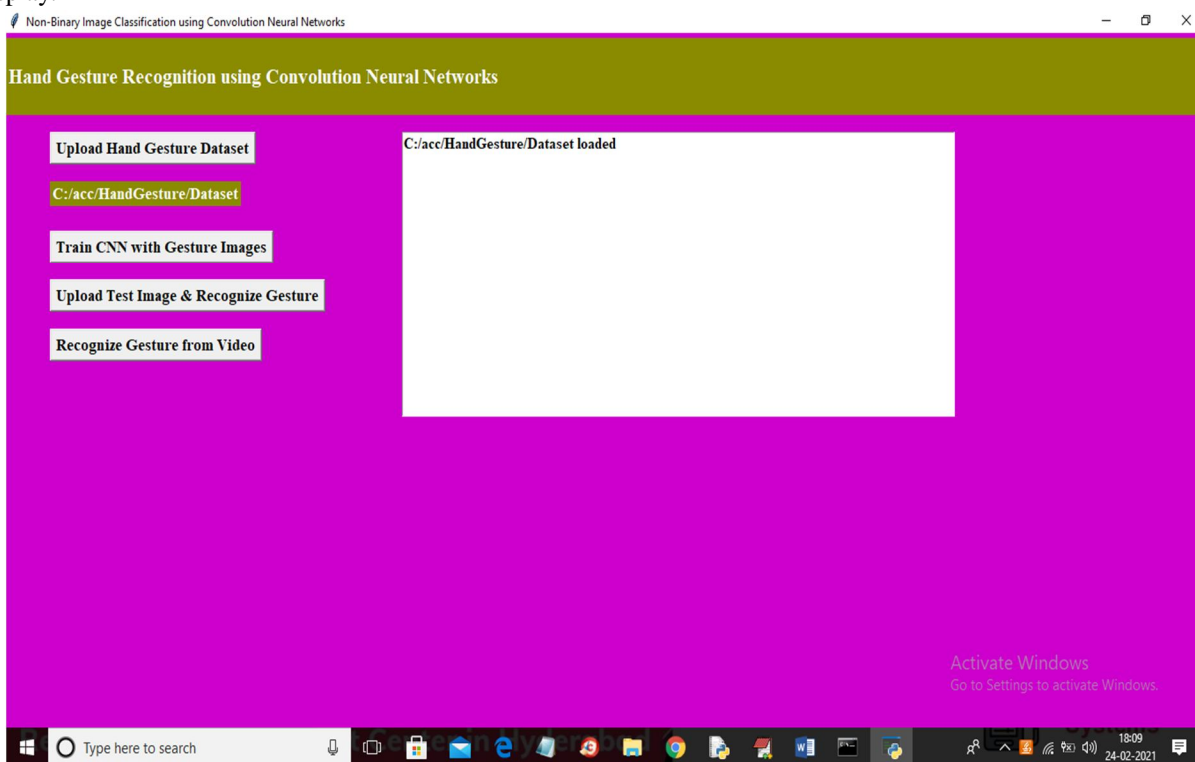
To run project double click on run.bat file to get below screen



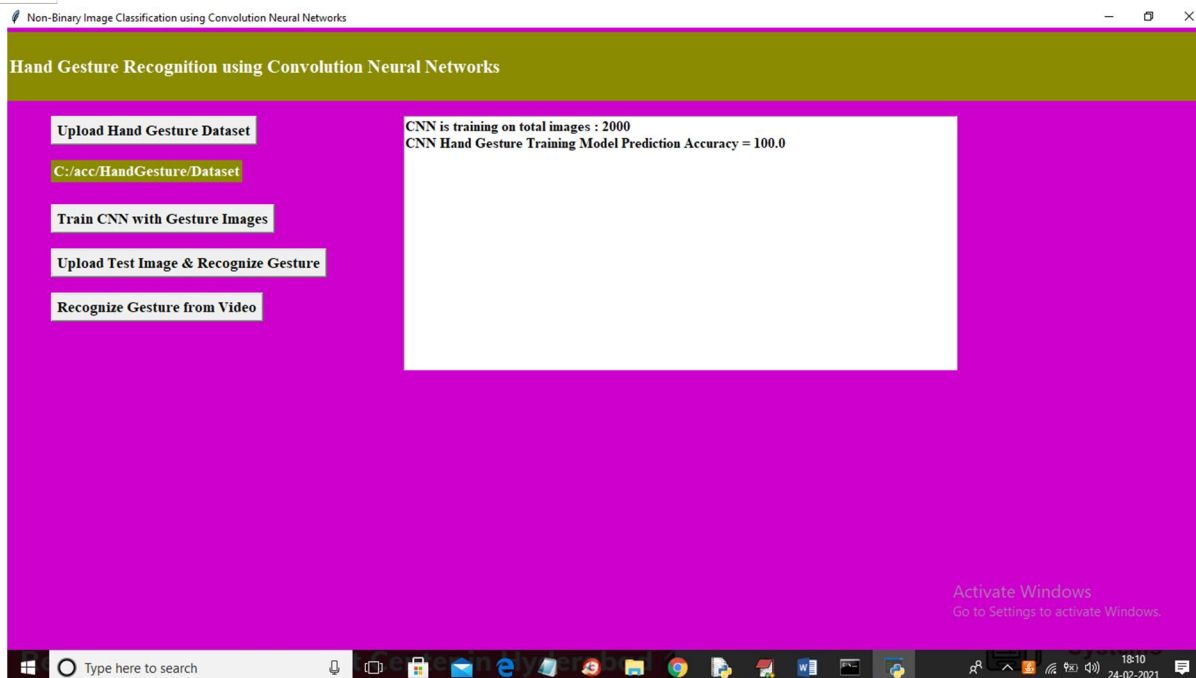
By clicking on the "Upload Hand Gesture Dataset" button that is located on the screen that came before this one, users will be able to reach the page that they are looking for, which will make the process of uploading the dataset easier.



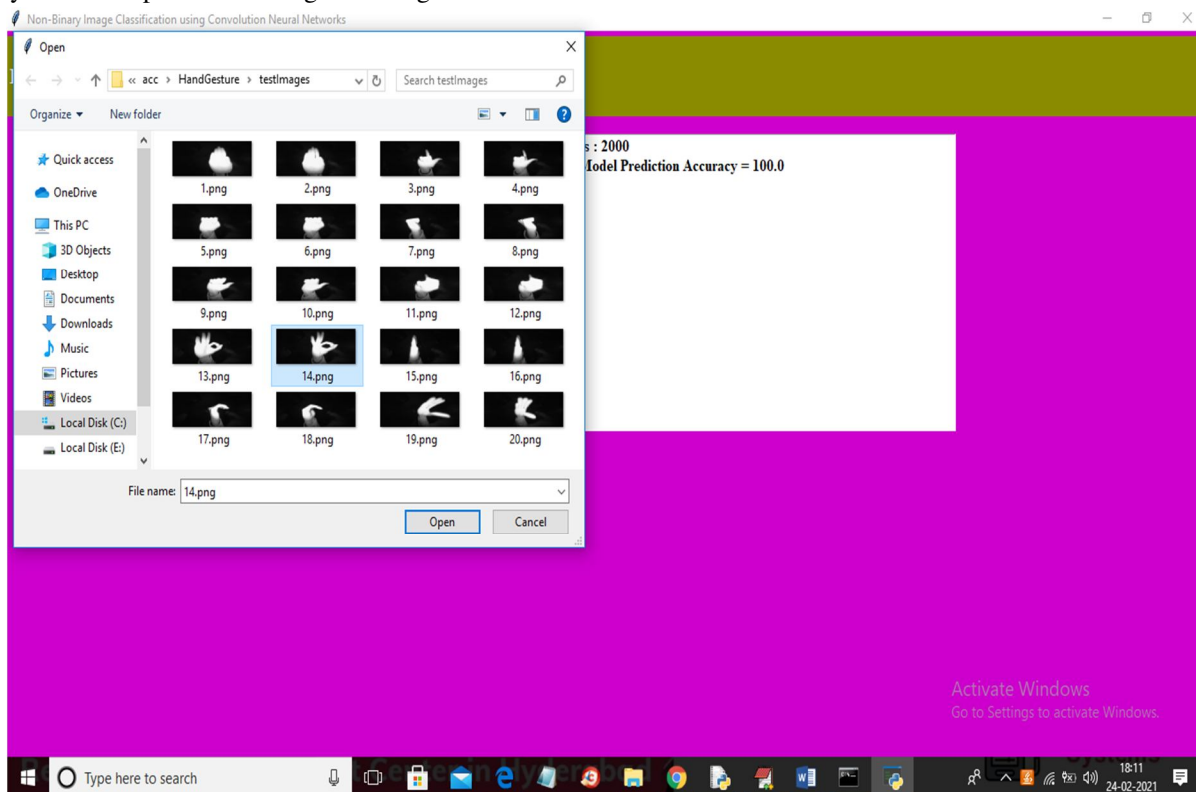
After that, the user will continue by choosing the 'Dataset' folder on the screen that was previously described and then uploading it. The user then proceeds to load the dataset by clicking the 'Select Folder' button, which allows them to proceed to the succeeding screen display.



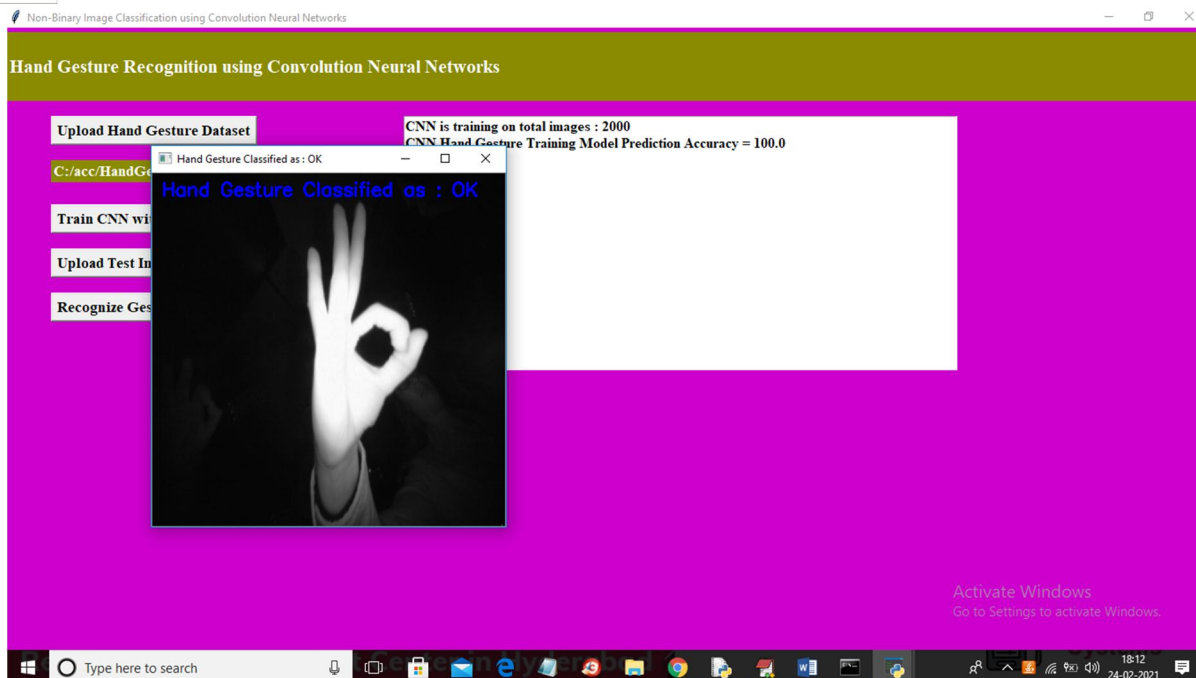
A successful import of the dataset into the display area located above has been accomplished. In order to start the process of training the Convolutional Neural Network (CNN) model and have access to the following presentation, kindly click on the button that is specifically marked as "Train CNN with Gesture Images."



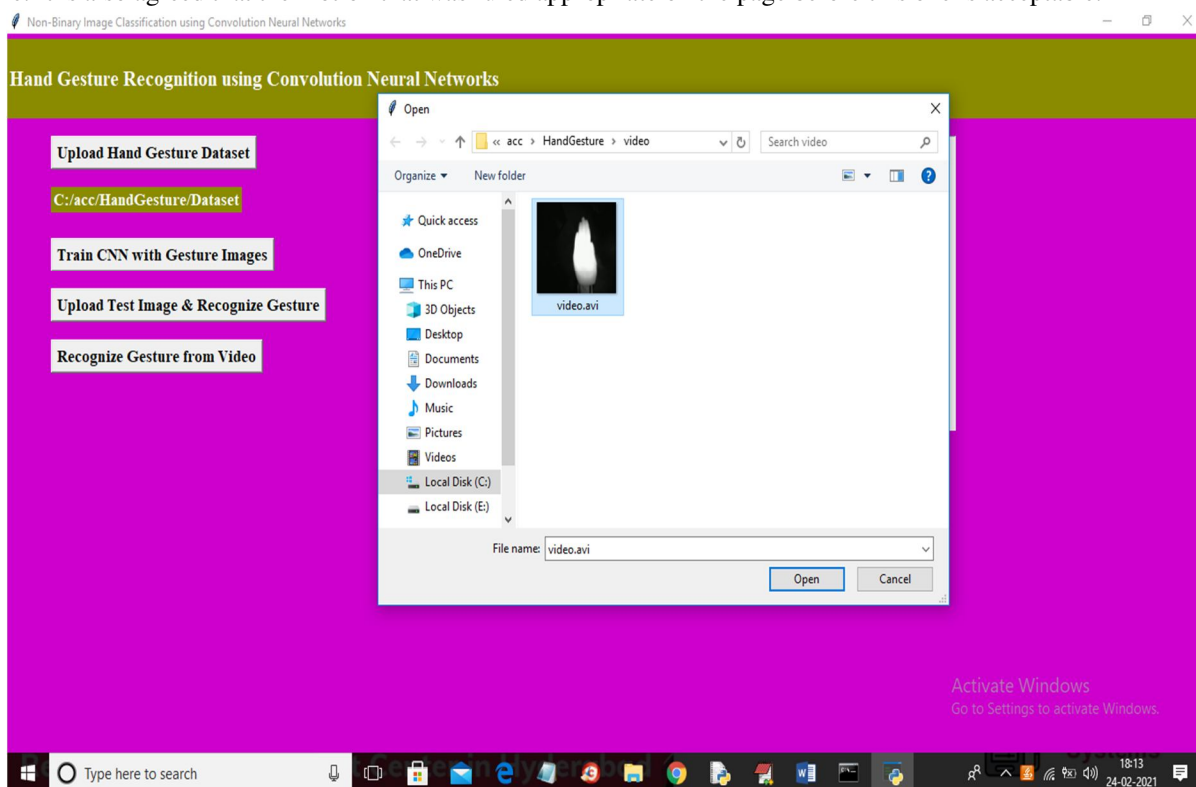
As can be seen in the display that is located above, the CNN model was trained using a dataset that had two thousand photos, and the accuracy of its predictions was verified to be one hundred percent. The model is now ready to be used in its current state. If you would like to begin the process of uploading a picture and then later recognizing movements, please choose the option that is specifically labeled "Upload Test Image & Recognize Gesture."



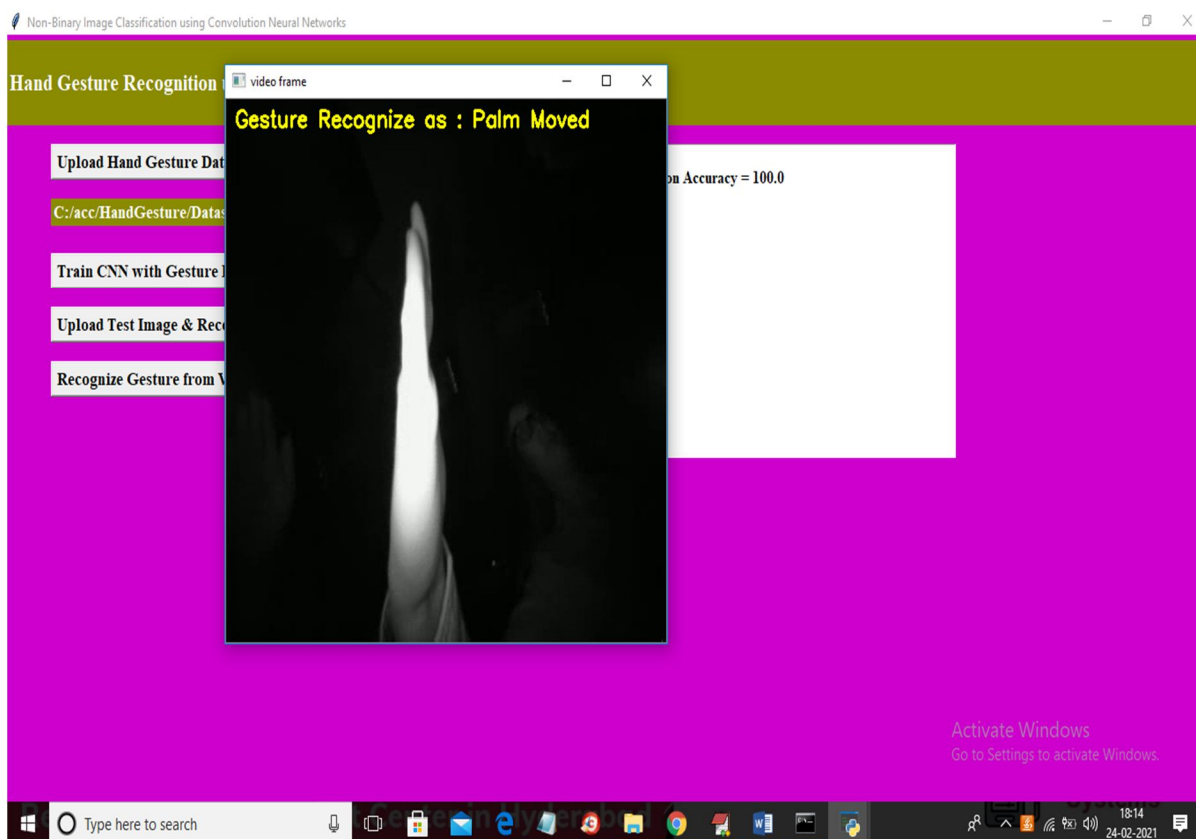
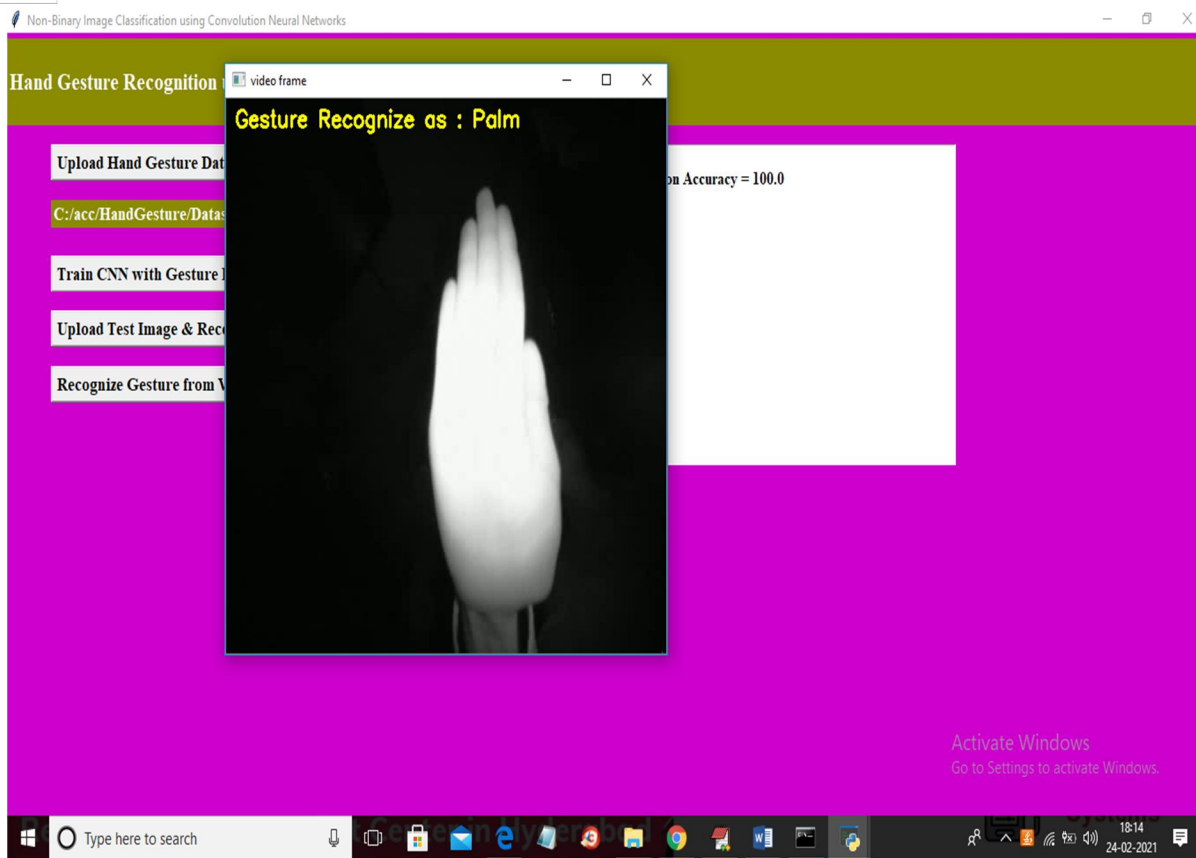
After selecting and uploading the file with the name "14.png" that was shown on the screen that was previously discussed and then choosing to hit the Open button, the subsequent result will be as described below.

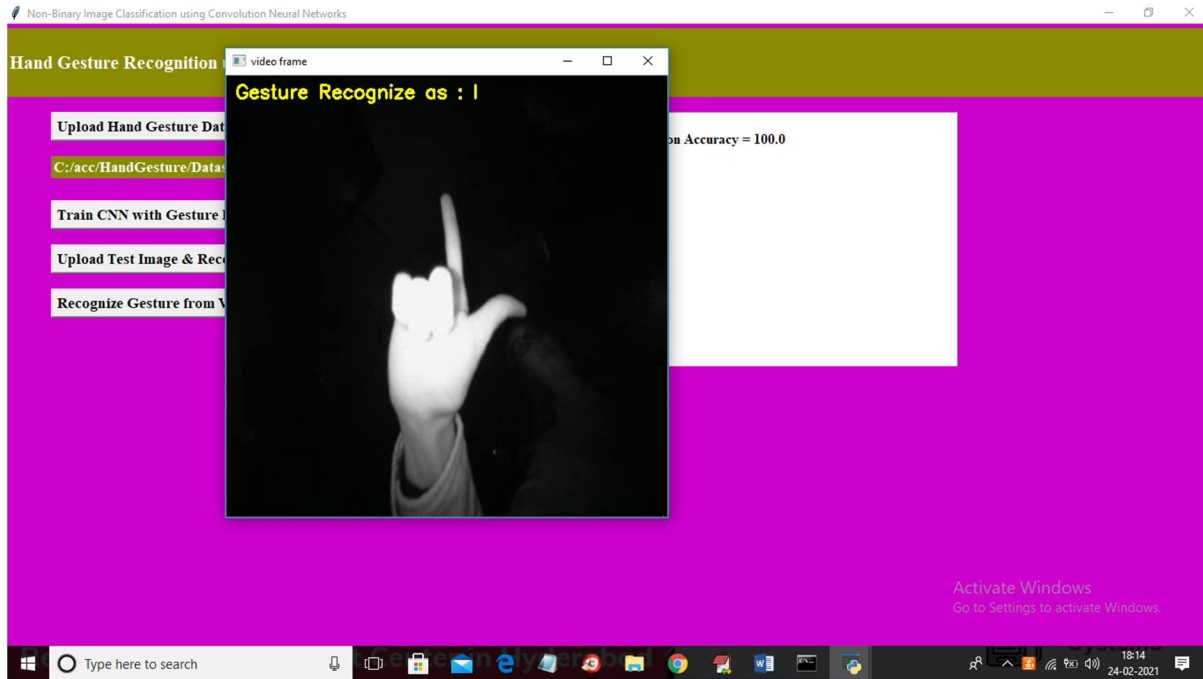


It is possible for users to upload a variety of photographs and get results that correlate to those images. Additionally, users now have the ability to submit films and acquire results by choosing the "Recognize Gesture from Video" button. This feature was previously unavailable. It is also agreed that the motion that was ruled appropriate on the page before this one is acceptable.



The user will need to click the 'Open' button in order to get the result that is shown below after they have chosen and uploaded the file with the name 'video.avi' on the page that was previously described.





In above screen as video play then will get recognition result

V. CONCLUSIONS & FUTURE SCOPE

The objective of the constructed CNN model was to facilitate the identification and interpretation of sign language through the use of inherent linguistic characteristics. The model demonstrates the capacity to acquire knowledge and extract spatially and temporally dimensional data by using three-dimensional convolutions. After extracting various types of information from successive input frames, the created deep architecture proceeds to execute convolution and subsampling using a novel strategy, different from the previously used method. The amalgamation of data from individual channels leads to the incorporation of information from each channel into the ultimate feature representation. A multilayer perceptron classifier, specifically used for the task of classification, achieves the classification of these feature representations. In order to provide a comparative analysis of the CNN and GMM-HMM models, it is important to subject both models to testing using an identical dataset. This article presents the experimental data, which demonstrate the efficacy of the proposed strategy.

REFERENCES

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [2] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei, "Large-scale video classification with convolutional neural networks," in CVPR, 2014. [3] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [3] Hueihan Jhuang, Thomas Serre, Lior Wolf, and Tomaso Poggio, "A biologically inspired system for action recognition," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. Ieee, 2007, pp. 1–8. [5] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3D convolutional neural networks for human action recognition," IEEE TPAMI, vol. 35, no. 1, pp. 221–231, 2013.
- [4] Kirsti Grobel and Marcell Assan, "Isolated sign language recognition using hidden markov models," in Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1997, vol. 1, pp. 162–167.
- [5] Thad Starner, Joshua Weaver, and Alex Pentland, "Realtime american sign language recognition using desk and wearable computer based video," IEEE TPAMI, vol. 20, no. 12, pp. 1371–1375, 1998.
- [6] Christian Vogler and Dimitris Metaxas, "Parallel hidden markov models for american sign language recognition," in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999, vol. 1, pp. 116–122.
- [7] Kouichi Murakami and Hitomi Taguchi, "Gesture recognition using recurrent neural networks," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1991, pp. 237–242.
- [8] Chung-Lin Huang and Wen-Yi Huang, "Sign language recognition using model-based tracking and a 3D hopfield neural network," Machine vision and applications, vol. 10, no. 5-6, pp. 292–307, 1998.
- [9] Jong-Sung Kim, Won Jang, and Zeungnam Bien, "A dynamic gesture recognition system for the korean sign language (ksl)," Systems, Man, and Cybernetics,



Part B: Cybernetics, IEEE Transactions on, vol. 26, no. 2, pp. 354–359, 1996.

- [10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” arXiv preprint arXiv:1311.2524, 2013.
- [11] Ronan Collobert and Jason Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in ICML. ACM, 2008, pp. 160–167.
- [12] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun, “Learning hierarchical features for scene labeling,” IEEE TPAMI, vol. 35, no. 8, pp. 1915–1929, 2013. [15] Srinivas C Turaga, Joseph F Murray, Viren Jain, Fabian Roth, Moritz Helmstaedter, Kevin Briggman, Winfried Denk, and H Sebastian Seung, “Convolutional networks can learn to generate affinity graphs for image segmentation,” Neural Computation, vol. 22, no. 2, pp. 511–538, 2010.
- [13] Ao Tang, Ke Lu, Yufei Wang, Jie Huang, and Houqiang Li, “A real-time hand posture recognition system using deep neural networks,” ACM Transactions on Intelligent Systems and Technology, 2014.
- [14] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt, “Sequential deep learning for human action recognition,” in Human Behavior Understanding, pp. 29–39. Springer, 2011.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)