

Analysis of Touchless Mouse Technology for Physical Disabilities

Renu Ahlawat

Research Scholar, Department of Computer Science and Engineering Baba Mastnath University, Asthal Bohar, Rohtak-124021

Dr. Banita

Associate Professor, Department of Computer Science and Engineering Baba Mastnath University, Asthal Bohar, Rohtak-124021

Abstract

We provide a touchless mouse system that surpasses past attempts by utilising deep learning models such as DenseNet169 and DenseNet201, in addition to an ensemble model. For feature extraction in the touchless mouse system, we use two different state-of-the-art convolutional neural network architectures, namely DenseNet169 and DenseNet201. These models, which were trained using massive datasets, perform remarkably well regarding computer vision tasks. Touchless mouse technology's sophisticated feature extraction capabilities make the exact recognition and interpretation of hand motions and movements possible. An ensemble model is developed by integrating the results of DenseNet169 and DenseNet201. This is done to make the system's performance even more effective. The ensemble technique improves the accuracy, stability, and generalizability of hand gesture detection by capitalising on these distinctions and using them to its advantage. Comparisons are made between the DenseNet169 and DenseNet201 models, the Ensemble model and several other deep learning and ensemble learning models. Additional deep learning and ensemble learning models are also displayed. The Ensemble model reached the maximum attainable accuracy of 99.62 per cent.

Keywords: DenseNet169, DenseNet201, Quadriplegia, Convolutional

Introduction

Physical disabilities, ranging from quadriplegia to severe motor impairments, present significant challenges to individuals in their daily lives, including the use of technology. In an increasingly digital world, access to computers and other electronic devices is essential for communication, education, and employment. However, conventional input methods, such as traditional computer mice and keyboards, can be inaccessible to many people with physical disabilities. In recent years, touchless mouse technology has emerged as a promising solution to address these accessibility barriers. This technology enables individuals to interact with computers and digital devices without physical contact, making it particularly valuable for those with limited or no motor control. By using alternative input methods such as gestures, eye-tracking, and brain-computer interfaces (BCIs), touchless mouse technology offers a pathway towards greater independence and inclusion for individuals with physical disabilities. This paper aims to comprehensively analyse touchless mouse technology for physical disabilities, exploring its various facets, including the types of touchless interfaces available, their usability in

real-world applications, associated challenges, and ethical considerations.

Research Methodology

Data Information

Sign language is used to collect data, found at <https://www.kaggle.com/datasets/muhammadrkhalid/sign-language-for-numbers>. It is a dataset for sign language specifically designed for number identification using hand gestures. Movements in sign language that correspond to clicks or other operations performed by the mouse; with the help of these datasets, the following gestures can be identified: There are ten different classes, including 0, 1, and 2. where two is considered to be a mouse click, 0 is considered to be a mouse drag, and other is considered to be movement.

Data Preprocessing

Touchless mouse data preprocessing is essential for efficient analysis and modelling. Some typical procedures for PR for use are: To begin, a sign language dataset is a dataset for sign language specifically designed for number identification

using hand gestures. Movements in sign language that correspond to clicks or other operations performed by the mouse. Filtering and wrapping methods are examples of feature selection strategies that can be used, such as data augmentation. Applying changes or perturbations to the original data is a popular data augmentation method used to expand the size and diversity of a dataset. Some data augmentation preparation techniques for a collection of mouse signs without touch:

- One translation is to make a minute adjustment to the hand or gesture in the recorded video frames or 3D data by moving it up, down, left, or right. This mimics the tiny shifts in hand posture that occur naturally when signing.
- Second, you can rotate the video frames or 3D data so that the hand or gesture faces a new direction. This aids in accurately representing the varying hand positions and motions used in sign language.
- Third, scale the hand or gesture in the video frames or 3D data. This takes into account any differences in hand size or camera-to-signer distance.
- Flipping: - Reverse the horizontal or vertical orientation of the video frames or 3D data. This improves the model's generalisation by providing more instances of sign language movements performed with either the left or right hand.
- Adjust the video's frames' brightness, contrast, and saturation to alter the scene's illumination. This considers the many settings where the touchless mouse mechanism may be employed.

Data augmentation strategies must be carefully calibrated to guarantee that the changed data accurately reflects real-world situations without introducing artificial or incomprehensible behaviour. For the sake of a fair evaluation of the model's efficacy, validation and testing data should not be inflated in any way.

Modelling

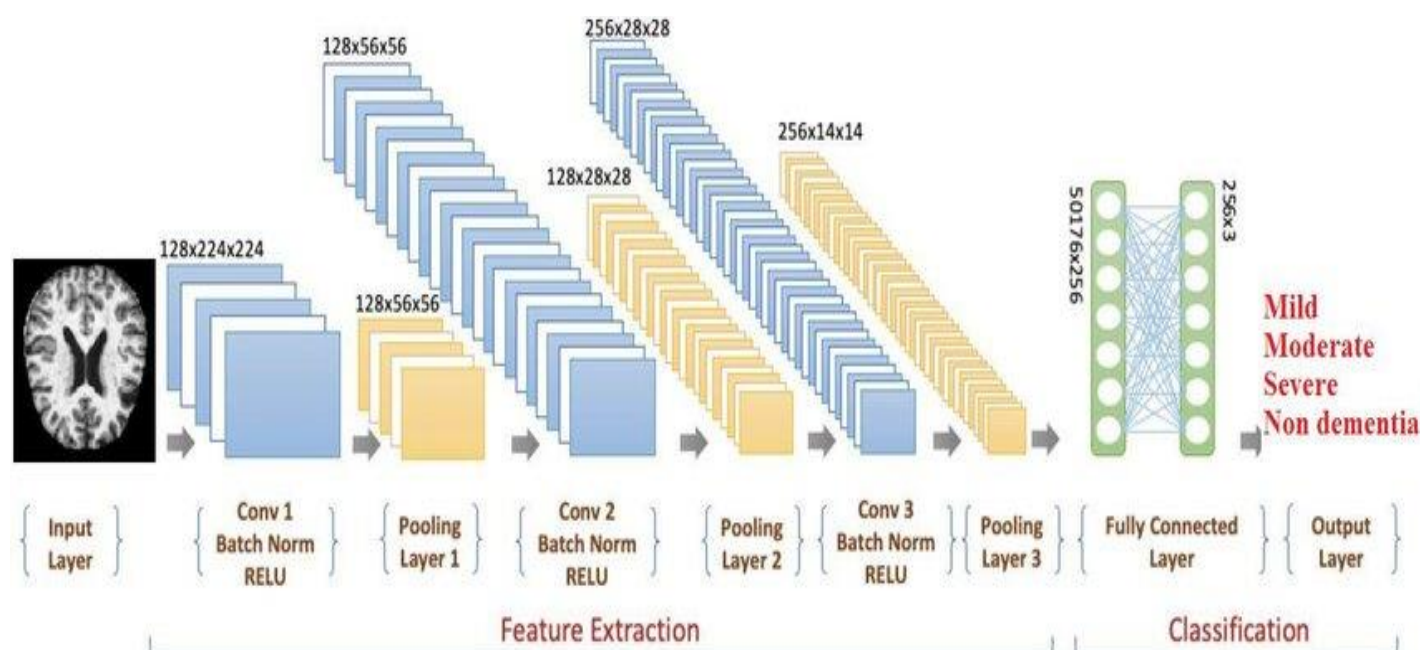
Deep Learning

Deep learning is an area of machine learning which concentrates on training artificial neural networks with many layers, also known as deep neural networks, to learn and generate predictions from complex data. Deep neural networks are sometimes referred to as multiple-layer artificial neural networks. Models trained with deep learning are especially useful for problems involving pattern recognition,

feature extraction, and hierarchical learning. Deep learning architectures such as DenseNet-169 and DenseNet-201 are components of the DenseNet family of models. Their extensive connectivity pattern, which enables efficient information flow and the reuse of features across several layers, is one of the defining characteristics of these networks.

- **DenseNet 169**

With its convolutional neural network (CNN) design, the DenseNet-169 model is one kind of CNN. The 2017 paper "Densely Connected Convolutional Networks" by Huang et al. was essential in setting the stage. DenseNet-169 is a variant of DenseNet that adds 169 layers. Here are a few ways in which DenseNet-169 stands out: High-Density Interactions: Each layer of DenseNet-169 is connected to every other layer in a feed-forward fashion, using the dense connectivity design. Much talk is happening between the layers, which is great for gradient flow, parameter efficiency, and feature propagation. Second, the dense blocks of DenseNet-169 have a "bottleneck" topology. - By limiting the number of input channels before applying the dense connections, the bottleneck structure conserves both memory and processing power. This architecture employs a 1x1 convolution layer followed by a 3x3 convolution layer to minimise the number of input channels. Thirdly, Compact Groups: The multiple thick layers of DenseNet-169 are organised into four separate dense blocks. Each dense layer inside a dense block is densely coupled to all other dense layers within the same block, meaning that each layer receives feature mappings from all previous layers within the same block as input, which helps to increase the model's depth and encourage feature reuse. Each pair of dense blocks in DenseNet-169 is linked together via transition layers. We can reduce the feature maps' width and height using transition layers while maintaining the same number of channels. Reducing the number of spatial dimensions helps keep the model's size and computational burden manageable. At the end of DenseNet-169, a global average pooling layer aggregates the spatial information inside the feature maps. Global average pooling averages down the spatial dimensions of feature maps to a single value, making classification and other post-processing stages easier. DenseNet-169 is particularly effective at classifying images thanks to its dense network and bottleneck design. After training on large-scale datasets like ImageNet, it achieves competitive accuracy with fewer parameters than previous CNN designs. DenseNet-169 is a foundation for further training or adaptation in various computer vision applications.



CNN Architecture for Brain MRI images Classification

Figure 1 DenseNet169 Architecture

• DenseNet201

2017's "Densely Connected Convolutional Networks" study by Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger presented the deep convolutional neural network architecture known as DenseNet-201. It is an upgrade to the original DenseNet design meant to increase gradient distribution and combat disappearing gradients. Each layer in DenseNet-201 gets input from all layers below it, creating dense connections between them. This dense connectivity topology boosts gradient propagation, feature extraction, and overall network performance by allowing for efficient information flow and feature reuse across different levels of the network. Some of the most notable aspects of the DenseNet-201 framework are several dense blocks: DenseNet-201 is made up of several dense blocks, each comprised of several convolutional layers of the same output size. The extensive connectivity between layers within a dense block encourages feature reuse and robust feature propagation. To limit the expansion of feature maps and lessen the computing burden, DenseNet-201 employs transition layers between dense blocks. By utilising pooling and convolution procedures, transition layers downsample the feature maps, lowering the spatial dimensions and the

number of feature maps. Adding more feature mappings to the output of each layer in a dense block is controlled by the growth rate, a hyperparameter. Each layer in the dense block of DenseNet-201 generates 32 more feature maps, for a total of 16384. Fourth, a fixed-length vector representation of the spatial features is generated by global average pooling, and then the network's output is classified. DenseNet201 is a deep convolutional neural network architecture that belongs to the family of DenseNets and is characterised by dense connectivity patterns. This architecture features 201 layers, utilising densely connected blocks to enhance information flow between layers. Dense connectivity involves concatenating feature maps from all preceding layers, fostering feature reuse and alleviating the vanishing gradient problem. DenseNet201 incorporates dense blocks with bottleneck layers, reducing computational complexity while maintaining model performance. The architecture employs global average pooling for spatial information aggregation and utilises batch normalisation and rectified linear unit (ReLU) activation functions for normalisation and non-linearity. DenseNet201 has demonstrated exceptional performance in image classification tasks, particularly on large-scale datasets, showcasing its ability to capture intricate patterns and representations across numerous layers.

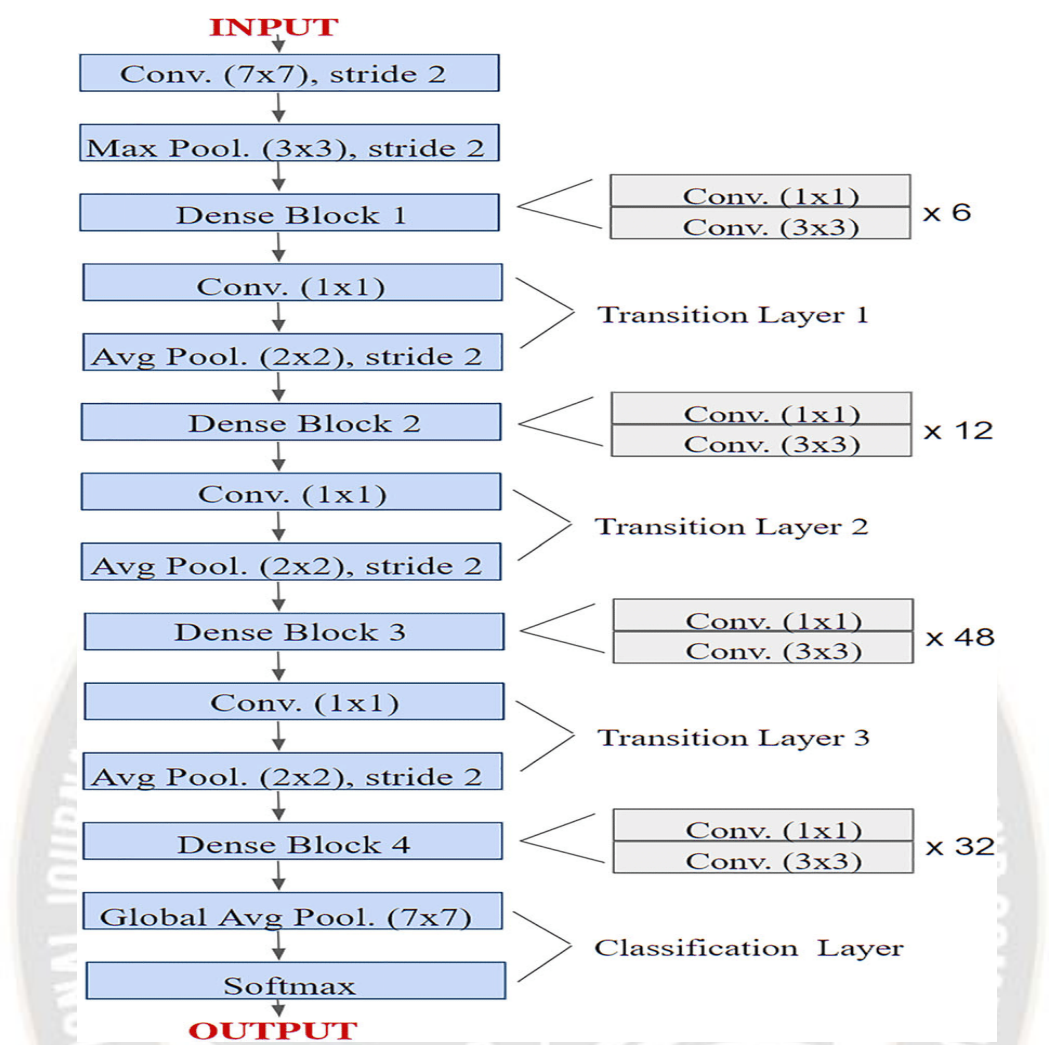


Figure 2 DenseNet201 Architecture

Results and Discussion

Two components make up the suggested solution. The first component is the training phase of a deep neural network, which generates outputs based on the input gauges. The second component is the application of a computer vision and image processing library to assign labels to every photo. Using the Keras library and Open CV for image processing, this study analyses and contrasts previously published and suggested ensemble learning models by putting into practice DenseNet169 and DenseNet201 into practice. Several alternative hyperparameters have been used to create the models in the table. These models include DenseNet201, DenseNet169, and an ensemble learning technique incorporating picture preprocessing. Both Open CV and the Keras library should be utilised when processing photos.

Hyperparameters Details

The table outlines the configuration details and specifications of a deep learning model designed for image classification, employing DenseNet169, DenseNet201, and an Ensemble model. The model is built for processing images with an input shape of 128x128x3 pixels. It utilises Conv2D layers with 'same' padding, followed by Global Average Pooling. Batch normalisation is incorporated for normalisation, and a dropout rate of 50% is applied to prevent overfitting. The activation functions used include Rectified Linear Unit (Relu) and Softmax. For performance evaluation, metrics such as Accuracy, Precision, Recall, and loss are employed. The model undergoes training for 100 epochs to learn and optimise its parameters. This comprehensive configuration captures the deep learning model's architectural aspects and training specifics, providing insights into the setup and evaluation criteria for image classification tasks.

Table 1: Deep Learning Specifications

| | |
|------------------------|--|
| Name | Deep Learning |
| Model | DenseNet169, DenseNet201, Ensemble model |
| Input Shape | 128*128*3 |
| Layers | Conv2D |
| Padding | Same |
| Pooling | Global Average Pooling |
| Normalisation | Batch |
| Dropout | 50% |
| Activation | Relu, Softmax |
| Loss | Categorical Cross Entropy |
| Performance Evaluation | Accuracy, Precision, Recall, loss |
| Epochs | 100 |

Relu Activation: The activation function Rectified Linear Unit (ReLU) is widely used in deep neural networks. Since it is a non-linear function, it enhances the network's expressiveness and capacity for pattern recognition and learning.

The ReLU activation function is defined as follows:

$$F(x) = \max(0, x) \quad (1)$$

Softmax: When dealing with data that falls into multiple categories, neural networks often use the softmax activation function in the output layer. The output is a probability distribution of the classes, while the input is a vector of real values. The softmax function prioritises classes with bigger values in determining the likelihood of each class given

$$Iw: softmax(x, i) = \exp(x, i) / \sum(\exp(x, j)) \quad (2)$$

The i-th element of the input vector x is denoted by x_i in this equation, and the total is calculated utilising all of the input vector elements using the exponential function denoted by exp(x_i). The softmax activation function is widely used in deep learning models for multi-class prediction problems like image classification, NLP, and audio identification.

Categorical Cross Entropy: The "categorical cross-entropy" loss function, also known as "softmax cross-entropy" or "cross-entropy loss," is frequently used in multi-class classification problems. Researchers developed the categorical cross-entropy loss function to quantify how far predictions deviate from reality. Since its major use is labelling input samples with a single class, it works best when the classes are distinct.

The formula available for the categorical cross-entropy loss is $CapL = -\sum(Y_{true} * \log(Y_{pred}))$ (3). The softmax activation function produces predicted class probabilities (y_pred), while y_true are the actual class labels (one-hot encoded). The sum is based on adding up all of the terms.

Performance Evaluation

The results of the study are shown here. Throughout the experiment, the feasibility of the proposed model was evaluated using a Python simulator and several performance measures. The authors used the Inception Resnet V2 and DenseNet models as transfer learning and ensemble learning strategies. The evaluation criteria are detailed below.

1) Accuracy

One measure of a classification model's effectiveness is its prediction accuracy. It counts the number of correctly labelled samples as a percentage of all samples. One way to evaluate the performance of a model is to look at how well it predicts outcomes for a specific group.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision

Precision is a performance metric that measures how accurately a test identifies positive samples. The elimination of unnecessary false positives is its primary focus. Precision is the percentage of correctly labelled positive samples by a model.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

2) Loss

The degree to which the model was wrong in its predictions is quantified in terms of loss in these other contexts. If a model fails to anticipate an outcome precisely, losses will rise, whereas a correct prediction will not affect the bottom line. Training models involve searching for an optimal combination of biases and weights to minimise the overall loss.

Recall

Recall that the fraction of properly predicted positive samples, as a percentage of the overall amount of positive samples, is used to evaluate performance. The eradication of false negatives is a current priority. Recall quantifies how well a model can catch and identify real positive samples properly.

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

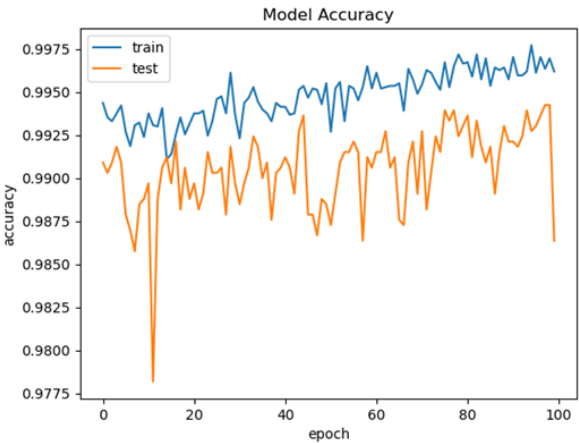
$$eCapLoss = -\frac{1}{m} \sum_{i=1}^m y_i \cdot \log(y_i) \tag{7}$$

Table 2: Performance evaluation of the models

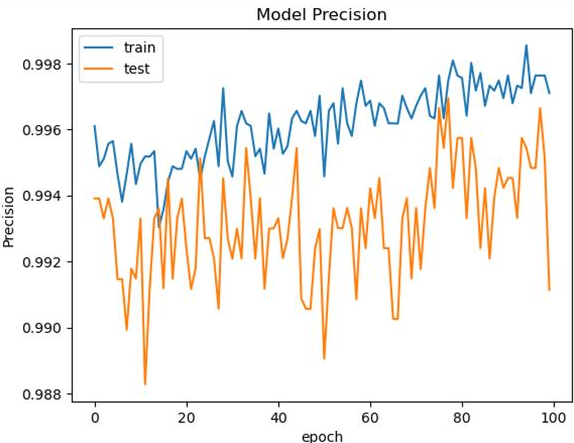
| Models | Accuracy | Precision | Recall | Loss |
|----------------|----------|-----------|--------|-------|
| DenseNet169 | 98.90 | 99.05 | 98.77 | 0.037 |
| DenseNet201 | 98.95 | 99.05 | 98.77 | 0.039 |
| Ensemble Model | 99.62 | 99.71 | 99.55 | 0.018 |

Table 1 provides a comprehensive overview of the performance evaluation of distinct deep learning and ensemble models, specifically DenseNet169, DenseNet201, and the Ensemble model. The Ensemble model emerges as the top performer, showcasing a maximum accuracy of 99.62%. Notably, when compared to individual models, the Ensemble model demonstrates superior performance. It outperforms DenseNet201 and DenseNet169 by achieving a 1% improvement in accuracy. This enhancement highlights

the efficacy of the Ensemble model in aggregating diverse model outputs to attain a more robust and accurate prediction. The results affirm the advantage of leveraging ensemble techniques in amalgamating the strengths of individual models, thereby contributing to heightened classification accuracy. The findings underscore the potential of ensemble learning as a strategic approach for enhancing the overall performance of deep learning models in the context of the studied task.



(a)



(b)

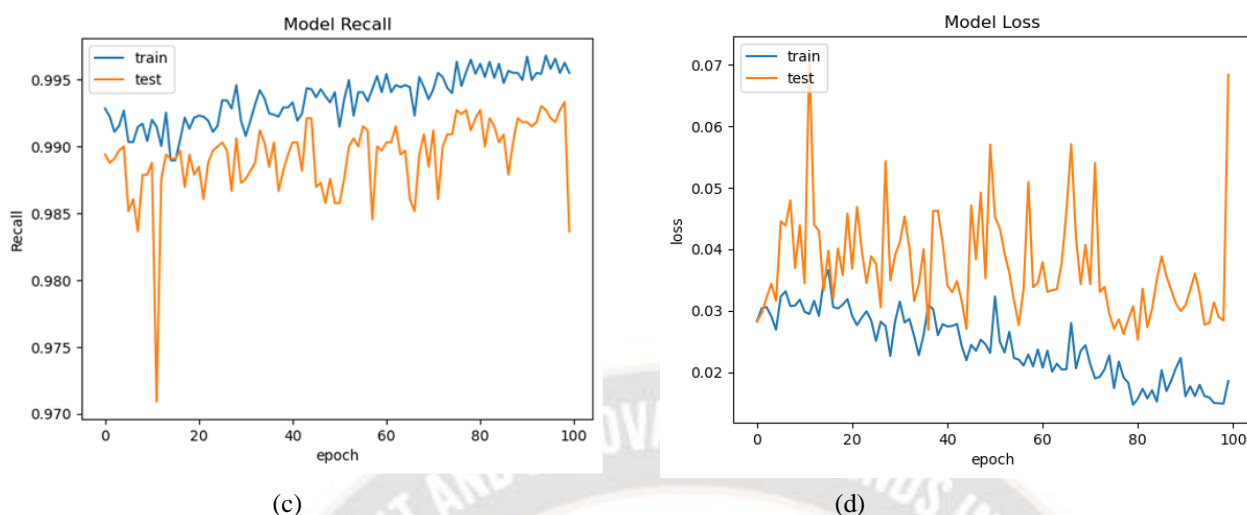


Figure 3: Results of Ensemble model (a) Accuracy (b) Precision (c) Recall (d) Loss

In Figure 3, a graphical representation illustrates the superior performance of the ensemble model compared to other models. The training phase is in blue, while the testing phase is in orange. The distinctive separation and dominance of the blue segment indicate the ensemble model's effectiveness

during training, surpassing the performance of alternative models. This visual representation underscores the ensemble model's capability to achieve superior results in training and testing scenarios, reaffirming its prowess in capturing complex patterns and enhancing overall model performance.

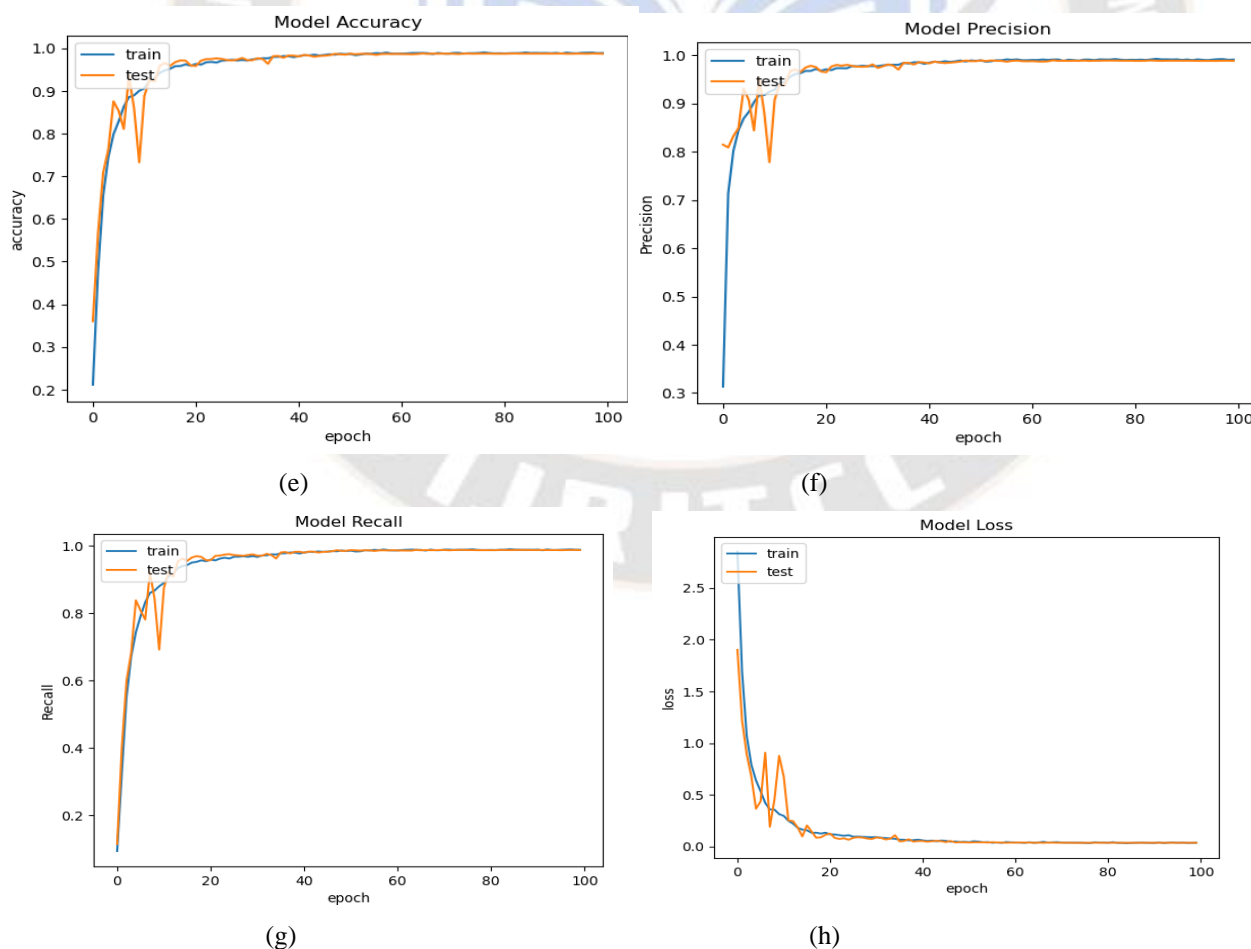


Figure 4: Results of DenseNet169 model (e) Accuracy (f) Precision (g) Recall (h) Loss

In Figure 4, a graphical depiction showcases the outcomes of the DenseNet169 model. The colour blue denotes the training phase, while the colour orange signifies the testing phase. This visual representation provides insights into the model's performance during training and testing. The distinct colour coding facilitates a clear interpretation of the model's

behaviour. It highlights its proficiency in learning from the training data (blue) and generalising on unseen test data (orange). This graphical representation is a valuable tool for assessing the model's training and testing dynamics, contributing to a comprehensive understanding of its overall performance.

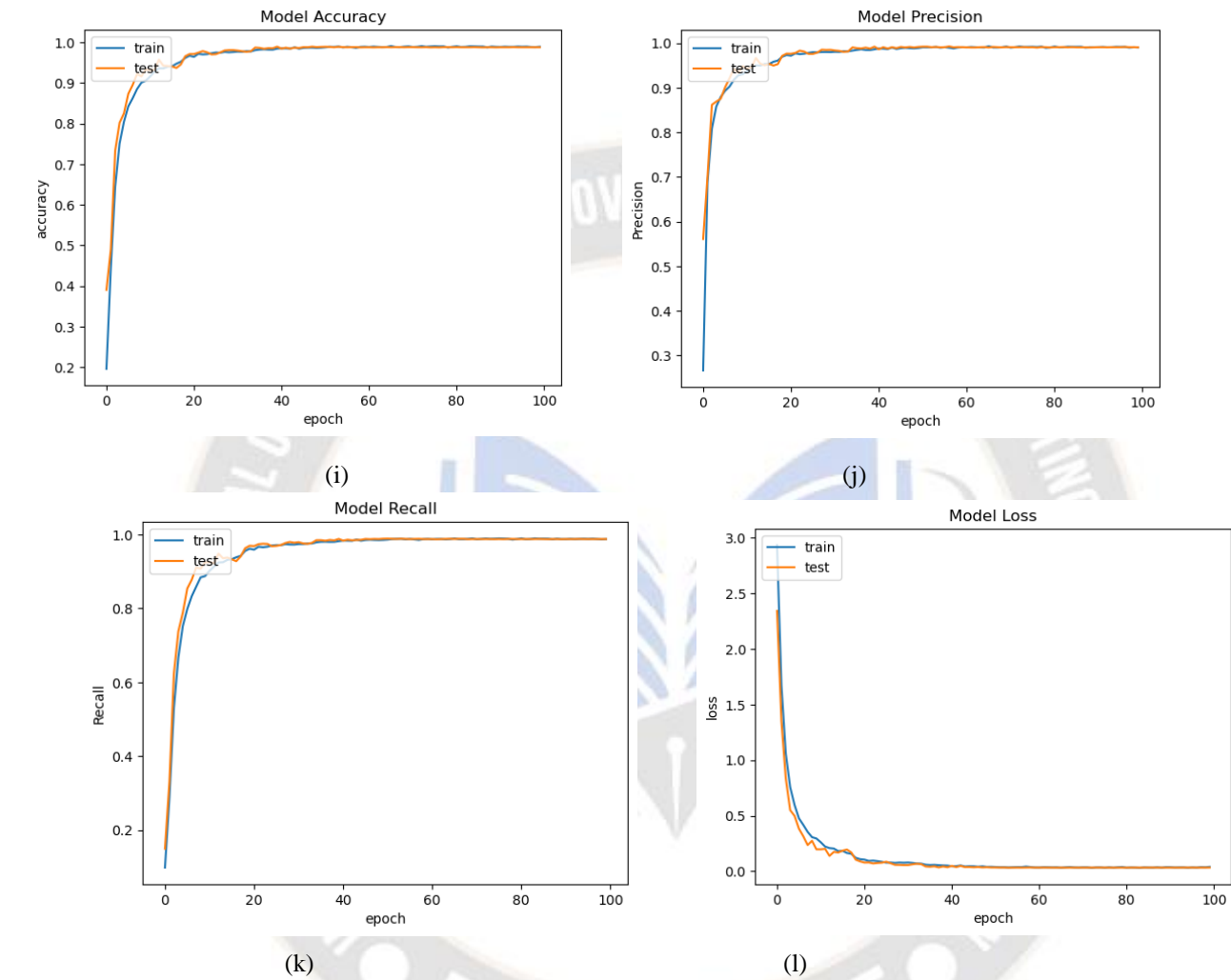


Figure 5: Results of DenseNet 201 model (i) Accuracy (j) Precision (k) Recall (l) Loss

In Figure 5, a graphical illustration delineates the outcomes generated by the DenseNet201 model. Blue indicates the training phase, symbolising the model's learning process. Conversely, orange is associated with the examination or testing phase, representing the model's performance on unseen data. This visual representation is a concise and

informative tool, offering a clear distinction between the training and testing dynamics of the DenseNet201 model. The colour-coded scheme visually dissects the model's behaviour during both phases, facilitating a nuanced assessment of its learning and generalisation capabilities.

Table 3: Comparative Analysis

| Model | results | references |
|-------------------------|---------|------------|
| GRU-based model | 96.78% | [15] |
| FaceNet model | 96% | [71] |
| Proposed Essemble model | 99.62% | -- |

The table outlines the performance metrics and configurations of various models employed for a task, possibly related to image classification or facial recognition. The GRU-based model achieved an accuracy of 96.78%, showcasing its capability to handle sequential data. The FaceNet model, known for facial recognition, attained an accuracy of 96%, demonstrating its effectiveness in

identifying faces. The proposed ensemble model, a combination of different models, outperformed both, achieving an impressive accuracy of 99.62%. The ensemble approach leverages the strengths of individual models, resulting in superior performance for the specific task at hand. These results emphasise the potential of ensemble models to enhance accuracy in complex classification scenarios.

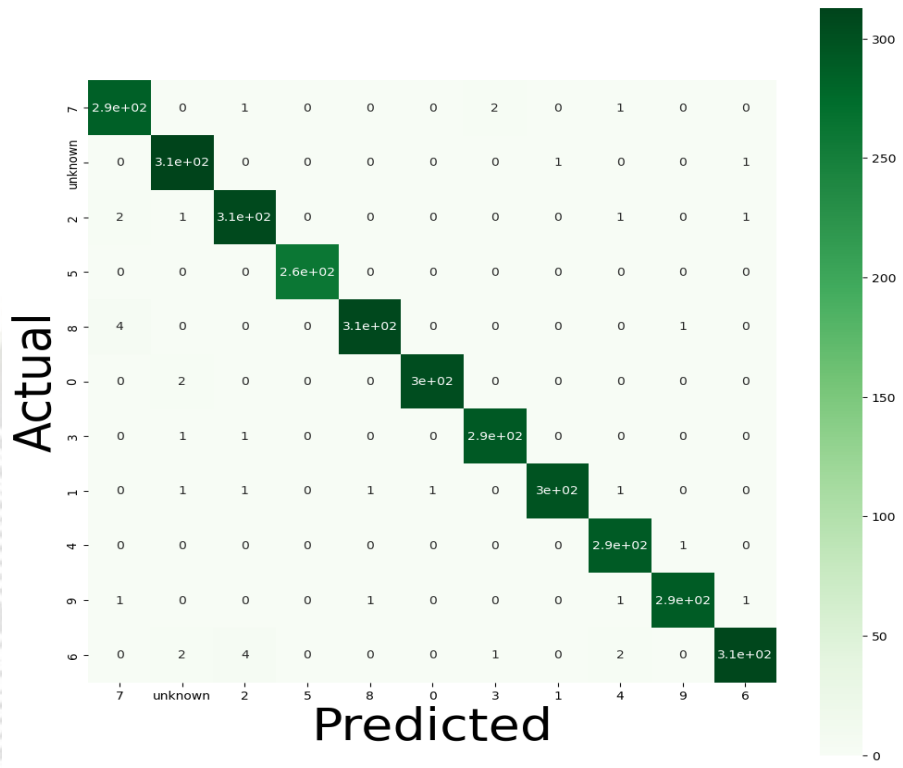


Figure 6 Confusion Matrix

In Figure 6, a confusion matrix displays both the actual values and the corresponding predicted values. This visual representation provides a comprehensive overview of the model's classification performance, allowing for a detailed **Conclusion**

The amalgamation of DenseNet169 and DenseNet201 within an ensemble model for touchless mouse technology represents a breakthrough beyond the mere amalgamation of neural network architectures. It signifies a strategic leap forward in enhancing the capabilities of touchless interfaces by harnessing the advanced feature extraction abilities of these deep convolutional neural networks. The efficiency demonstrated by DenseNet169 and DenseNet201 in various computer vision tasks has laid the foundation for their integration into touchless mouse technology. Their prowess in handling intricate visual data has been instrumental in significantly improving the precision and reliability of the touchless mouse system. The touchless mouse can accurately

examination of true positives, true negatives, false positives, and false negatives, aiding in assessing model accuracy and effectiveness.

discern and interpret hand movements and gestures by leveraging these models' enhanced feature extraction capabilities. This means that users can experience a more intuitive and responsive touchless mouse interface, as the technology is now equipped to interpret and respond to a broader range of hand gestures. The ensemble model, which strategically combines the strengths of DenseNet169 and DenseNet201, has proven particularly effective, achieving an impressive accuracy rate of 99.62%. This not only outperforms individual models but also demonstrates the synergistic benefits of combining complementary neural network architectures.

Moreover, the ensemble model's 1% improvement over the individual DenseNet201 and DenseNet169 models highlights

its ability to minimise error rates further. This enhancement is crucial in ensuring a seamless user experience, as it reduces misinterpretation and enhances the overall reliability of the touchless mouse system. As we reflect on the implications of this research, it becomes evident that integrating DenseNet169, DenseNet201, and ensemble techniques goes beyond immediate applications. It paves the way for future breakthroughs in touchless interface technology. The success achieved in performance and accuracy solidifies the current state of touchless mouse technology and sets the stage for continued innovation and refinement. Strategically incorporating DenseNet169, DenseNet201, and ensemble techniques has propelled touchless mouse technology to new heights. The resulting improvements in efficiency, precision, and user experience underscore the transformative impact of this research, making significant contributions to the evolution of human-computer interaction. As we navigate the ever-expanding landscape of technology, the successful integration of these advanced neural network models serves as a beacon, guiding future developments in touchless interfaces and beyond.

References

- Chahin, N. Krishnan, H. Chhatrala, and M. Shaikh, "A 5-Fluorouracil-Induced Hyperammonemic Encephalopathy Challenged with Capecitabine," *Case Rep. Oncol. Med.*, vol. 2020, pp. 1–4, 2020, doi: 10.1155/2020/4216752.
- S. Akram *et al.*, "Construction and Analysis of a Novel Wearable Assistive Device for a Visually Impaired Person," *Appl. Bionics Biomech.*, vol. 2020, 2020, doi: 10.1155/2020/6153128.
- Skrabaneck, P. Dolezel, Z. Nemec, and D. Stursa, "Person Detection for an Orthogonally Placed Monocular Camera," *J. Adv. Transp.*, vol. 2020, 2020, doi: 10.1155/2020/8843113.
- B. Šumak, M. Špindler, M. Debeljak, M. Heričko, and M. Pušnik, "An empirical evaluation of a hands-free computer interaction for users with motor disabilities," *J. Biomed. Inform.*, vol. 96, no. March 2018, p. 103249, 2019, doi: 10.1016/j.jbi.2019.103249.
- S. F. Paulo *et al.*, "Touchless interaction with medical images based on 3D hand cursors supported by single-foot input: A case study in dentistry," *J. Biomed. Inform.*, vol. 100, no. September, p. 103316, 2019, doi: 10.1016/j.jbi.2019.103316.
- Y. Komano, K. Ohta, K. Sakiyama, M. Iwamoto, I. Verbauwhede, and D. Schneider, "Single-round pattern matching key generation using the physically unclonable function," *Secur. Commun. Networks*, vol. 2019, no. Id, 2019, doi: 10.1155/2019/1719585.
- T. Hu, W. Niu, X. Zhang, X. Liu, J. Lu, and Y. Liu, "An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning," *Secur. Commun. Networks*, vol. 2019, 2019, doi: 10.1155/2019/3898951.
- Samian, A. H. Zaidan, Patricia, R. N. Afifah, and M. Yasin, "Touchless Mechanism to Detect Rhodamine B Concentration in Distilled Water Using Fiber Bundle," *Int. J. Opt.*, vol. 2019, 2019, doi: 10.1155/2019/5918958.
- F. Bouabdallah, "Time Evolution of Underwater Sensor Networks Coverage and Connectivity Using Physically Based Mobility Model," *Wirel. Commun. Mob. Comput.*, vol. 2019, 2019, doi: 10.1155/2019/9818931.
- E. R. M. Aleluya and C. T. Vicente, "Faceture ID: Face and hand gesture multi-factor authentication using deep learning," *Procedia Comput. Sci.*, vol. 135, pp. 147–154, 2018, doi: 10.1016/j.procs.2018.08.160.
- H. S. Grif and T. Turc, "Human hand gesture-based system for mouse cursor control," *Procedia Manuf.*, vol. 22, pp. 1038–1042, 2018, doi: 10.1016/j.promfg.2018.03.147.
- M. Nasor, K. K. M. Rahman, M. M. Zubair, H. Ansari, and F. Mohamed, "Eye-controlled mouse cursor for a physically disabled individual," *2018 Adv. Sci. Eng. Technol. Int. Conf. ASET 2018*, no. February, pp. 1–4, 2018, doi: 10.1109/ICASET.2018.8376907.
- D. S. Lopes *et al.*, "On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface," *J. Biomed. Inform.*, vol. 72, pp. 140–149, 2017, doi: 10.1016/j.jbi.2017.07.009.