information

Article

A Novel Deep Transfer Learning Approach Based on Depth-Wise Separable CNN for Human Posture Detection

Roseline Oluwaseun Ogundokun [1], Rytis Maskeliūnas [1], Sanjay Misra [2,*] and Robertas Damasevicius [1]

[1] Faculty of Informatics Engineering, Kaunas University of Technology, 51368 Kaunas, Lithuania
[2] Department of Computer Science and Communication, Østfold University College, 1757 Halden, Norway
* Correspondence: ssopam@gmail.com

Abstract: Human posture classification (HPC) is the process of identifying a human pose from a still image or moving image that was recorded by a digicam. This makes it easier to keep a record of people's postures, which is helpful for many things. The intricate surroundings that are depicted in the image, such as occlusion and the camera view angle, make HPC a difficult process. Consequently, the development of a reliable HPC system is essential. This study proposes the "DeneSVM", an innovative deep transfer learning-based classification model that pulls characteristics from image datasets to detect and classify human postures. The paradigm is intended to classify the four primary postures of lying, bending, sitting, and standing. These positions are classes of sitting, bending, lying, and standing. The Silhouettes for Human Posture Recognition dataset has been used to train, validate, test, and analyze the suggested model. The DeneSVM model attained the highest test precision (94.72%), validation accuracy (93.79%) and training accuracy (97.06%). When the efficiency of the suggested model was validated using the testing dataset, it too had a good accuracy of 95%.

Keywords: deep transfer learning; human posture classification; silhouettes; human posture

Citation: Ogundokun, R.O.; Maskeliūnas, R.; Misra, S.; Damasevicius, R. A Novel Deep Transfer Learning Approach Based on Depth-Wise Separable CNN for Human Posture Detection. *Information* **2022**, *13*, 520. https://doi.org/10.3390/info13110520

Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 26 September 2022
Accepted: 28 October 2022
Published: 31 October 2022

1. Introduction

The volume of imagery and video resources is growing at an incredible rate in the era of internet media. The computer vision (CV) industry has much room to use data to develop techniques that can create improved monitoring systems by confirming human posture, thus preventing the occurrence of ailments such as cardiovascular disease. By using a person's posture history or activities as a guide, one may forecast their posture using the human posture (HP) categorization technique. In essence, the notion is that once the subject's activity is identified and recognized, an intelligent computer system model can offer support. Deep learning (DL) algorithms have been successful in recent studies for human posture detection (HPD) [1,2].

The goal of HPD is to identify poses using still images or video sequences. The development of medical systems, the detection of improper postures, autonomous technologies, and scenario assessment are just a few areas where image-based posture classifications are crucial. Taking into account some elements such as complicated backgrounds in videos or images, interactions between individuals that cause body parts to overlap, occlusions, and other difficulties, HPC is a difficult process. The sole function in the detection and classification of human postures is a 2D human pose estimation (HPE). Approaches to estimating human poses take into account information on human posture [1].

Due to the ability of DL techniques to effectively perform the training task when presented with large data sets, it is particularly effective in handling the majority of computer vision problems [3–9]. The main form of the DL technique is deep transfer learning (DTL), and these techniques are used to solve a variety of issues such as object recognition, face recognition systems, image classification, and image detection. In this study, four DTL approaches are used to develop an approach of classifying human posture.

In the research article [10], a combined approach to predict human position that combines CNN and graph models was used. A two-stage system is shown, comprising two modeling techniques: one is a multistage CNN approach made to collect the necessary features from the imagery, and the next is a forecasting technique that projects the posture using the information recovered by the CNN technique [11].

Contribution

A sedentary lifestyle harms the human body and poor posture can cause neck, back, and shoulder discomfort if ignored. Poor posture can also lead to serious diseases, such as cardiovascular diseases. The study describes three significant contributions that are described as follows:

1. The study implemented novel hybrid DenseNet121 and SVM techniques to automatically recognize human postures.
2. The suggested model used regularization, early stop, and dropout techniques for L1 and L2 to prevent overfitting.
3. The layers of the DenseNet121 deep transfer learning (DTL) model were fine-tuned to achieve better results.
4. A comparative analysis of the outcomes implemented, and the existing system was conducted.

The main aim of this study is to examine the ability of the DeneSVM model to recognize human posture. The DTL technique used in this study is to improve the performance of human posture recognition. The standing, sitting, bending, and laying positions can be recognized. The ability of humans to monitor their postures for an extended time makes posture vital to recognize.

The work is divided into many sections, Section 2 provides a brief overview of techniques for classifying human posture. The approach for the suggested model for categorizing human posture is then presented in Section 3. The results are demonstrated in Section 4, and the conclusions and recommendations for future improvement are provided in Section 5.

2. Related Works

A survey of the research has been conducted to understand the approach and restrictions of many HPE techniques that are now in use. Table 1 provides a summary of the research.

The analysis of previous approaches revealed that strong algorithms are required since human posture estimation still faces difficulties. Therefore, this work suggested a deep transfer learning-based human posture detection approach, which is presented in Section 3.

Table 1. Summary of HPC pieces of literature.

Authors	Approaches	Contributions	Limitations
Du et al. [12]	This study presents a technique for identifying human activity by employing a transfer-learned residual network based on microdoppler spectrograms.	The authors attained an accuracy of 97% Jitter, which is less and the results deflection is less.	Their study lacks the computational capacity to train the data set used in this study.
Shi et al. [13]	The authors used DL approaches on spectrogram data. The adversarial generative network employed a generative adversarial network and DCNN	The study offers good scalability and uses less time for the training process	They obtained an accuracy of 82% and the study has high Jitter. Their results are also deflected.

Table 1. *Cont.*

Authors	Approaches	Contributions	Limitations
Cao et al. [3]	An effective technique is suggested to recognize 2D poses from several images comprising manifold persons. The technique used was the greedy bottom-up analysis technique.	The study delivers high accuracy with a low Jitter. The outcomes are very unreflective and suitable for instantaneous utilization.	The study requires extra computation power
Ning et al. [7]	This paper suggests a technique to intensify the accuracy of categorization by enhancing the single-shot detection (SSD) technique. The technique is developed by integrating architectural features	The technique used is vigorous.	The jitter is too high; the result frequently deflects with a lower accuracy for small images.
Caba et al. [10]	Three probable applications where ActivityNet can be employed were suggested. These include untrimmed video categorization, trimmed activity categorization, and activity recognition.	Additional varieties of classification and activity variety	The accuracy attained is low
Sung et al. [14]	A sensor that is not expensive, called RGBD was used in the input dataset. The authors employed a two-layer maximum entropy Markov model (MEMM).	The study achieved a training accuracy of 84.3%	The research achieved a low accuracy of 64.2% on the testing dataset
Karpathy et al. [15]	The authors used the UCF-101 video dataset, and this led to the development of a slow fusion model using CNN.	The sluggish fusion model was performed on early and late fusion networks, and it was demonstrated that the system had an accuracy of 80% on UCF-101.	There was little or no difference in the human image interaction, the individual-individual body-motion interaction, and the playing devices.
Laptev et al. [16]	Local space-time features, space-time pyramids, and nonlinear SVM of the manifold channel were employed to enhance existing outcomes on the standard KTH action dataset	The suggested technique attains a precision of 91.8% on the KTH action dataset	The suggested technique requires an enhancement in script-to-video gathering to a much larger dataset.

3. Materials and Methods

Computer vision and the detection of human postures are now both incredibly productive fields of study. An input and output layer for classification and several hidden layers make up a standard DTL model. Convolutional layers, pooling layers, fully connected layers (FC) and, in certain circumstances, SoftMax layers make up the hidden layers of a CNN. Most CNN architectures adhere to the Lenet-5 architectural design pattern of LeCun et al. [17]. There are several architectural designs already implemented. As a result, our work evaluated current DTL approaches and optimized them to classify and detect human posture using images from the Kaggle collection [18]. There are three architectures under consideration: InceptionV3, ResNet-50V2, and DenseNet (121 layers). To ensure that appropriate steps can be taken early to avoid complications and later mortality, rapid and precise systems for human posture detection are sought.

3.1. Dataset

DTL approaches were assessed and trained on images of human posture to identify and classify human posture on image positions that the suggested approach had not perceived earlier. The study used silhouettes for the human posture detection (HPD) dataset [18]. The data set was combined to recognize human posture. The dataset includes four postures: lying, bending, sitting, and standing. The data set is 1200 images for each of the postures, as shown in Figure 1. The image dimension is 512 × 512 pixels and the

distribution of the data set is given in Table 2. The data set is divided into four classes as seen in Table 3. The link to the data set is https://www.kaggle.com/datasets/deepshah16/silhouettes-of-human-posture accessed on 4 August 2022 [18].

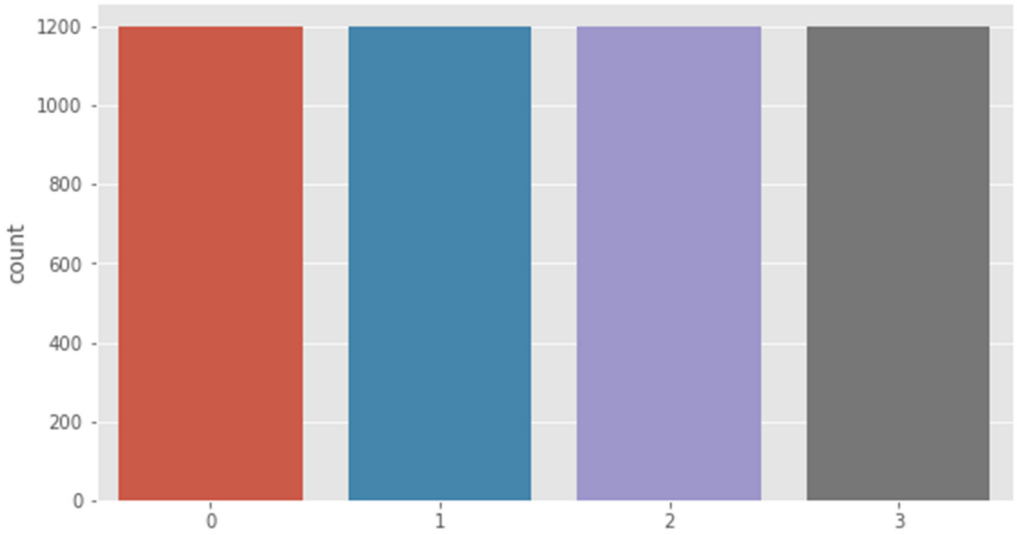


Figure 1. Human posture recognition for the four data sets (0 = bending, 1 = lying, 2 = sitting, and 3 = standing).

Table 2. Data Split.

Posture Classes	Training Set	Validation Set	Testing Set
0	867	153	180
1	867	153	180
2	867	153	180
3	867	153	180
Total	**3468**	**612**	**720**

Table 3. Overview of Silhouettes for HPD and Class Distribution.

Class Label	Posture Class Name	Number of Postures
0	Bending	1200
1	Lying	1200
2	Sitting	1200
3	Standing	1200
Total		**4800**

Since the data sets used in this study did not provide directives on distributing the data set in the train, validation, and test sets, the study utilized 70:15:15; that is, 70% of the data set was used for the training set (TRS), 15% for the validation set (VAS), and 15% for the testing set (TES) as shown in Table 2. As shown in Table 2, 15% of the initial data set was split, which is 720 given the remaining data set to be 4080; thereafter, 15% of the remaining 4080 data set was split as the validation set, which is 612, and then the remaining 3468 which is 70%, was used as the training set. The data test set is used for the detection and evaluation of the suggested approach.

3.2. Data Preprocessing

This represents the initial stage. The goal of data preprocessing is to convert the raw photos into a format that will be more practical and efficient to use in subsequent processing phases. The categorical method was used to normalize the photos in the initial stage. As a result of all training images having the same scale, such as between 0 and 1, normalization might save training time. The second pre-processing technique employed by the authors is data augmentation, which involves adding copies of existing data that have been slightly altered. As a regularizer, it decreases overfitting during the model training phase. The third way involves applying the least absolute shrinkage and selection operator (LASSO) feature selection methodology to create a weight vector that represents the relative relevance of the feature groups. The feature groups with high weights are chosen because they are considered more significant.

3.3. Existing DTL Images Models

In this section, deep-transfer learning models such as densenet121, resnet50v2, and inceptionv3 approaches are discussed. In this section, the researchers also present the proposed hybridized DeneSVM model (a combination of DenseNet121 and SVM). The fine-tuning of the proposed model is also discussed.

3.3.1. Suggested Approach

Figure 2 shows the suggested system flow using the silhouettes for human posture recognition dataset obtained from the Kaggle repository. The link to the data set is https://www.kaggle.com/datasets/deepshah16/silhouettes-of-human-posture accessed on 4 August 2022.

In this study, the use of fine-tuned hybridized DenseNet-121 and SVM called the "DeneSVM Model" is proposed and is evaluated against existing conventional DTL models such as InceptionV3, ResNet-50V2 and DenseNet-121. The images are initially colored images of diverse dimensions. The images were first resized to 244 × 224 pixels for all models (DeneSVM, InceptionV3, ResNet-50V2 and DenseNet-121). To make the network data consistent with the model prediction, normalization was carried out by dividing all pixel values by 255. Additionally, a hot encoding of the images was conducted to be able to use them for the model training.

3.3.2. InceptionV3

Szegedy et al. [19] were the ones who originally included the idea of "Inception" in the GoogleLeNet architecture. Inception vN, where N is the version number, was used to designate the subsequent iterations of the GoogleLeNet architecture. Updates to the Inception module are suggested in the publication by Szegedy et al. [19] to increase the accuracy of ImageNet classification accuracy. This is known as the Inception V3 architecture.

The pooling layer and convolution layers that make up the Inception module are placed on top of each other. The convolutions come in a range of sizes, including 1 × 1, 3 × 3, and 5 × 5. The Inception module's usage of a bottleneck layer, a 1 × 1 convolution, is another salient characteristic. Reduced computing needs are made possible by the bottleneck layer. Moreover, the module's pooling layer is utilized to reduce the dimensions of objects. A concatenation filter is necessary to combine layers, as demonstrated by Szegedy et al. [19].

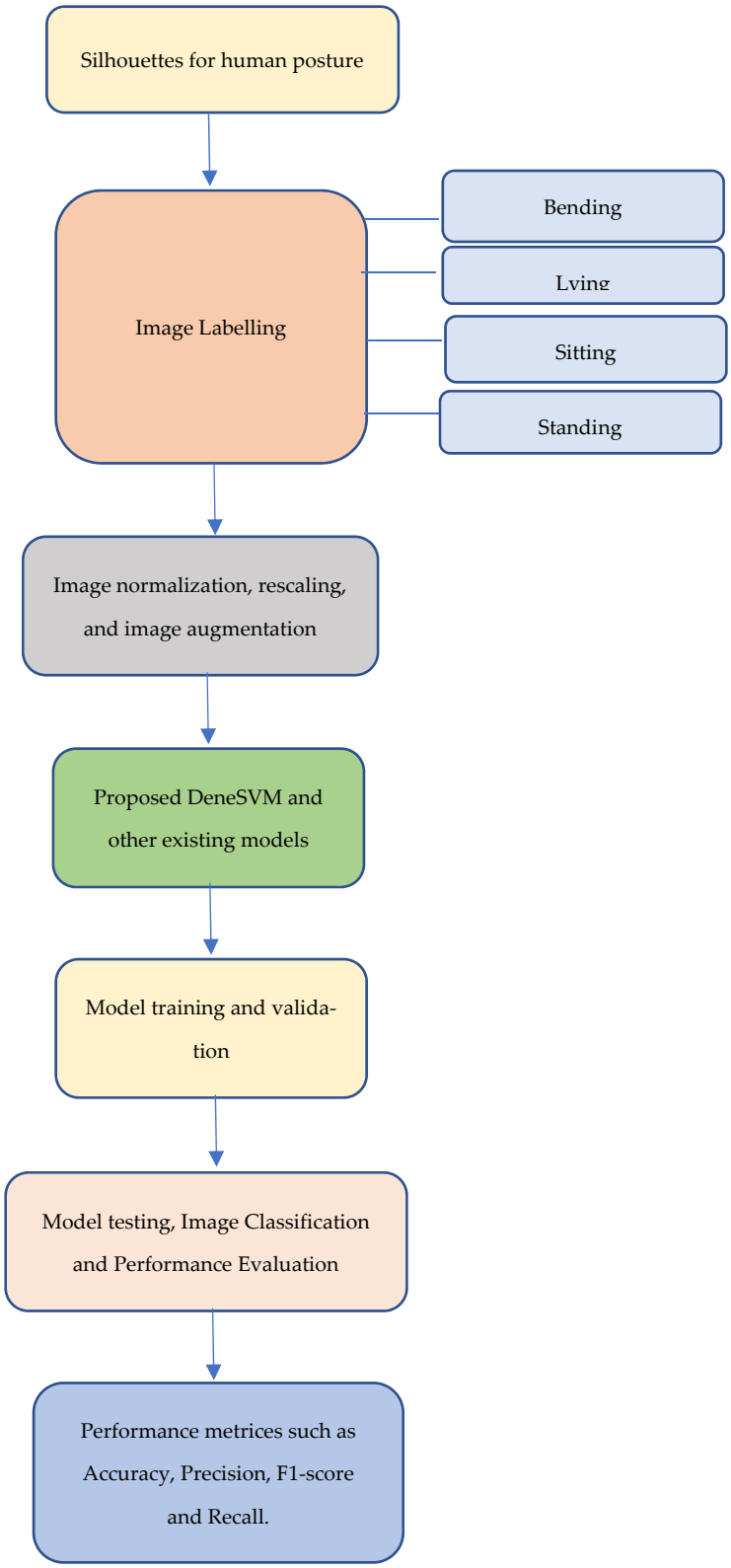


Figure 2. The proposed system flow using the Human Posture Images.

3.3.3. ResNet

The ResNet network, which served as the categorization challenges of foundation for the ILSVRC and COCO in 2015, was established by He et al. [20]. With an error rate of 3.57%, their method took the first place in the ImageNet classification. The deep residual

learning approach was prompted by the deterioration issue and the difficulty of numerous non-linear layers to acquire identity mappings (ResNet).

A network-in-network (NIN) system called ResNet relies on several layered residual modules. The network was established using this collection of residual modules as its structural pieces. The ResNet structure is made up of pieces of construction made up of residual modules [20]. The VGG net [21], which consists of 3×3 filters, is equivalent to residual modules in that it has convolution, pooling, and LS; however, ResNet is approximately eight times deeper than the VGG network. The use of global average pooling, as opposed to completely linked layers, is blamed for this. ResNet [22] was updated once again to improve accuracy using identity mappings in the residual layer.

The ResNet network with 50 version 2 was created and loaded with pretrained weights from ImageNet. The model was fine-tuned and, lastly, a modified SoftMax layer was built for the activity of human posture classification.

3.3.4. DenseNet

Huang et al. [19] developed a highly linked CNN model. Every tier in the network is intrinsically linked to every other layer in a feedforward approach to achieve a high amount of information flow between the levels. The extracted features are used as entries for each layer, and the specific extracted features are used as inputs for all following layers. The problem can be solved with DenseNets, which also significantly decreases the number of parameters [23].

Integration of DenseNet-121 and SVM networks was used to generate the DenseNets model with 121 layers, as reported by Huang et al. [23], to classify and detect human postures. The approach also has pre-trained weights from ImageNet injected into it. Eventually, a further FC model was produced, this time using the study's unique SoftMax as the top layer.

The purpose of SVM was to introduce non-linearities into the structural model by using kernel functions, which can improve the prediction capacity by allowing the entry of large amounts of data into a small space of feature space [24]. The study applied the radial basis kernel function to the SVM classifier of the DeneSVM model.

3.4. Fine-Tuning the Models

A transfer learning (TL) idea is fine-tuning (TL). TL is an ML methodology in which training from one kind of challenge is used to train in another similar task or method [25]. In DL, the initial layers are trained to recognize task-specific characteristics. Researchers dropped the final few layers of the classification algorithm during TL and retrained with new agents for the desired task. Although they take some time to perfect the learning process, fine-tuned learning experiments are still much faster than starting from the beginning [26]. Furthermore, they appear to be significantly more intelligent than models that were created from scratch.

The suggested DeneSVM TL model was improved to detect and classify four kinds of human posture using pre-trained architectures on the ImageNet dataset to speed up training procedures. There are 1000 class categories and around 1.2 million photos in the ImageNet collection. However, the human posture dataset is inadequate to build deep convolutional networks such as TL networks, which is why pre-trained parameters from ImageNet are used instead. With the addition of image enhancement, CNN DenseNet-121 was fine-tuned using an SVM classifier on the HPD dataset. The approach was developed and fed with ImageNet pretrain parameters. Furthermore, a custom softmax was defined on the top layer to truncate the top layer. Moreover, the Adam method and an initial learning rate of 0.001 were used to fine-tune the DeneSVM model and the other standard TL CNN models. The three architectures are listed along with their parameters in Table 4.

Table 4. Models and their specifications.

Model	Layers	Parameters	Layers in Based Model	Size (MB)	Depth
InceptionV3	48	23.9M	311	92	189
DenseNet-121	121	8.1M	427	33	242
ResNet-50V2	164	25.6M	190	98	103

The study used numerous existing DTL models established a few years ago, such as residual networks, densely connected networks, convolutional neural networks, and frameworks that facilitate model examination. The study implemented algorithms that were pre-trained (TL) on ImageNet, a sizable dataset of around 1.5 million realistic scenery images classified into 1000 classes, to satisfy the DTLs' insatiable desire for the dataset. The authors improved these algorithms using the advantages of transfer learning in silhouettes for HPD datasets. From the innumerable deep CNN architectures that have been used in the past, the authors eventually selected InceptionV3, ResNet50V2, and DenseNet-121 for their good performance. By integrating the densely linked network DenseNet-121 with the SVM ML technique, the researchers were able to merge the capabilities of each of these neural networks with biological inspiration. The study only fine-tuned the proposed DeneSVM model. For the proposed DeneSVM, the sum of layers in the base model was 427, and to fine-tune the proposed DeneSVM model, the first 404 layers were frozen, and the remaining was 23 layers.

3.5. Model Uncertainty

Practical machine learning has given significant attention to deep learning approaches. Unfortunately, these regressions and classifications do not consider the uncertainty of the model. On the contrary, Bayesian models have high processing costs and provide a theoretically valid framework to analyze the uncertainty of the model [27]. The effectiveness cannot be simply measured as accuracy at the software level. So, the researchers added a new indicator: the confidence score, to show how confident the proposed model is in our detection and categorization. An excellent way to measure uncertainty is to use a confidence score. The suggested model is a non-Bayesian network. The two main non-Bayesian methodologies for estimating uncertainty are Monte Carlo dropout (MC dropout) [28] and Deep Ensembles [29]. The dropout approach is one of the most popular strategies to prevent overfitting. Gal et al. [28] demonstrate that the uncertainty of the model can be determined by choosing the Bernoulli distribution with probability, such as a dropout probability. During the training and testing phases of MC Dropout, the dropout layer is utilized, and numerous predictions are then produced on a single image to determine the degree of uncertainty. For this study, MC Dropout was chosen. It requires fewer processing resources and smaller hyperparameters [30].

Since the dropout layer was introduced to the model after each fully connected (FC) layer, there are only two dropout layers in the study. To avoid overfitting, the dropout layer is frequently used during the training process. To ensure consistency in the prediction outcome for the same image, the dropout will be carried out during the analysis process. In MC, the dropout layer must be turned on. Prediction phase dropout results in a shift in the softmax value for each prediction, which impacts how it is categorized. Then, for each target image, 100 predictions are made, with the majority of the results from the estimations acting as the classification for the following projections. The number of forecasts will determine what percentage of the score represents confidence. No positive or negative forecast will be higher than 80 times in a sample of 100 predictions if the confidence score value is less than a predefined threshold, such as 80%. The researchers think that this situation is challenging to foresee and necessitates precise manual processing.

4. Results

The subsections can be utilized to subdivide this section. It should contain a clear and succinct explanation of the research observations, their assessment, and any possible experimental inferences.

4.1. Implementation

This section involves reporting the execution of the proposed DeneSVM model and baseline models such as InceptionV3, ResNet50V2, and DenseNet121. The configuration of the system used for the implementation is also presented here. The section presents similarly all the tasks that took place during the training phase.

4.1.1. Implementation Setup

The base system utilized in the proposed system is the workstation CPU, OpenCV, Keras, and Panda, which are employed for the software execution of the proposed model. The study used Python programming language for implementation on a Windows 11 system. The configuration of the system is as follows: Core i7, 16Gb RAM, and 1Terabye SSD hard disk.

4.1.2. Training

The suggested approach is evaluated for each execution using the accuracy measure and categorical cross-entropy loss (loss). All models' accuracies and losses when it is implemented are visually shown. The effectiveness of the recommended DeneSVM is evaluated using actual loss and accuracy values based on the test data set. Table 5 presents the findings obtained from the experimental processes carried out. All experiments include an early stopping technique and run for a total of 30 epochs. The epoch is the total number of training cycles. The decision was made to use 30 epochs to see which models could converge quickly and which models have deteriorating issues.

Table 5. Accuracy and loss of training, validation, and testing.

Model	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)	Training Loss	Validation Loss	Testing Loss
InceptionV3	80.07	90.20	92.08	0.4436	0.3824	0.3361
ResNet-50V2	85.47	89.54	91.53	0.3518	0.7053	0.3784
DenseNet-121	91.06	91.99	93.47	0.2420	**0.3242**	**0.2421**
DeneSVM	**97.06**	**93.79**	**94.72**	**0.1318**	0.4338	0.2918

In each of the techniques implemented, the hyperparameters were harmonized. Adam optimizer is used to build all networks, since it executes quickly and converges easily. Due to CPU memory limitations, researchers trained the networks using a 32-batch size. For all networks, the learning rate was set at 0.001. In all implementations, the study used the Relu activation function [31], and image enhancement was performed for each of the modeling techniques.

4.2. Results of Implementation

The deep transfer learning (DTL) model has been trained on the silhouettes for human posture recognition dataset, which comprises positions corresponding to bending, lying, sitting, and standing. Although the data set is already balanced, the study still made use of image data augmentation to increase the training dataset. The model was executed on a Dell laptop with a Windows 10 working platform. The setup of the system is an Intel Core i7, with 16GB RAM. The model is trained using TensorFlow with a batch size of 32 and a learning rate of 0.001. The model is trained using a categorical cross-entropy loss function and an increasing number of epochs using the Adam optimizer.

Various DTL models that had already been created over the years were used, such as the inception network, the residual network, densely linked network, and the frameworks for model exploration. The authors' employed algorithms that were pre-trained (TL) on ImageNet, a sizable dataset of around 1.5 million realistic scenery images subdivided into 1000 classes, to satisfy the DTLs' insatiable desire for the dataset. On Silhouettes for HPD datasets, the study improved these classifiers to make use of TL. Due to the strong performance of TL, the authors chose the DTL architectures InceptionV3, ResNet-50V2, and DenseNet-121. To present the findings, the study hybridized the densely connected network DenseNet-121 with the SVM ML technique, combining the capability of each of these biologically inspired computational models.

For training, the images were resized to 224 by 224 pixels before feeding them to the proposed DTL models. The data augmentation was then carried out with posture images randomly flipped horizontally with a flip probability of 0.2, a rotation range of 20 degrees, a width shift range of 0.3, a height shift range of 0.3, a shear range of 30, and a zoom range of 0.2. Hybridization and inference are taken to deliver the concluding results. The initial learning rate (LR) is set to 0.001. It is monitored using the validation accuracy, LR patience set is 10, verbose is 1, factor is 0.70 and the min_lr set is 0.00001. This indicates that whenever there is no improvement in the validation accuracy, the LR is reduced, especially after 10 epochs. To avoid the model from overfitting, the study introduced the early stopping technique, and it was set as thus: verbose is 1 and patience is 20. Two regularization techniques were introduced to avoid the model from overfitting and increase the performance accuracy. The networks are trained for 30 epochs and the number of training epochs, and the initial learning rate were decided analytically.

This research is an evaluation of the suitability of the existing DTL techniques for the task of identifying human postures using human position. Our main objective was to fine-tune the hybridized DenseNet-121 and SVM. DTL architectures are trained and fine-tuned following the instructions in Section 3.4. Figures 3–6 show the results of the experiment. Each graphic shows each model's precision and entropy log-loss, including the suggested DeneSVM model. After 10 epochs, the DeneSVM model had a training accuracy of more than 92%, InceptionV3 had an accuracy of more than 74%, ResNet-50V2 had more than 76% and DenseNet-121 had an accuracy of more than 86%, although the validation precision was over 93% for DeneSVM model, over 87% forInceptionV3, over 85% for ResNet-50V2, and over 90% for DenseNet-121. Additionally, extremely accurate values were achieved for both DeneSVM and InceptionV3, with significantly lower log-loss even after the 30th training iteration.

The proposed DeneSVM, InceptionV3, and DenseNet-121 models perform consistently better than ResNet-50V2. Furthermore, they effortlessly converge, as perceived in Figure 3. As seen in Table 5 and the confusion matrix (Figure 4), the hybridized approach performed better on the test dataset. Fewer iterations are required for satisfactory performance from DeneSVM, InceptionV3, and DenseNet-121. ResNet-50V2, on the contrary, underperforms with fewer iterations, as seen in Figure 3. However, ResNet-50V2 improves its accuracy and reduces its log loss with more iterations, as seen in Figure 3b.

Overall, the ResNet-50V2 model performed poorly with the least accuracy and the largest log loss, as shown in Table 5 and Figure 3b, while the suggested Novel DeneSVM model suggested performed effectively with the best accuracy and the least log loss, as shown in Figure 3d.

The test data set was run and utilized to assess the models and the confusion matrix was generated for each of the models. The proposed model along with the existing DTL models is shown in Figure 4. Table 6 shows the results obtained from the implementation of the test data set to assess the performance of the suggested system using performance measures such as precision, recall and f1-score. Table 6 revealed that using the testing dataset (dataset not seen by the model or used for training and validation) for performance evaluation, the proposed DeneSVM model outperformed with an accuracy of 95% when likened to the other existing models. InceptionV3 had the lowest accuracy of 92%, which is

also a good model performance to some extent. The proposed DeneSVM had the best AUC curve value of 99.36%, as shown in Table 6.

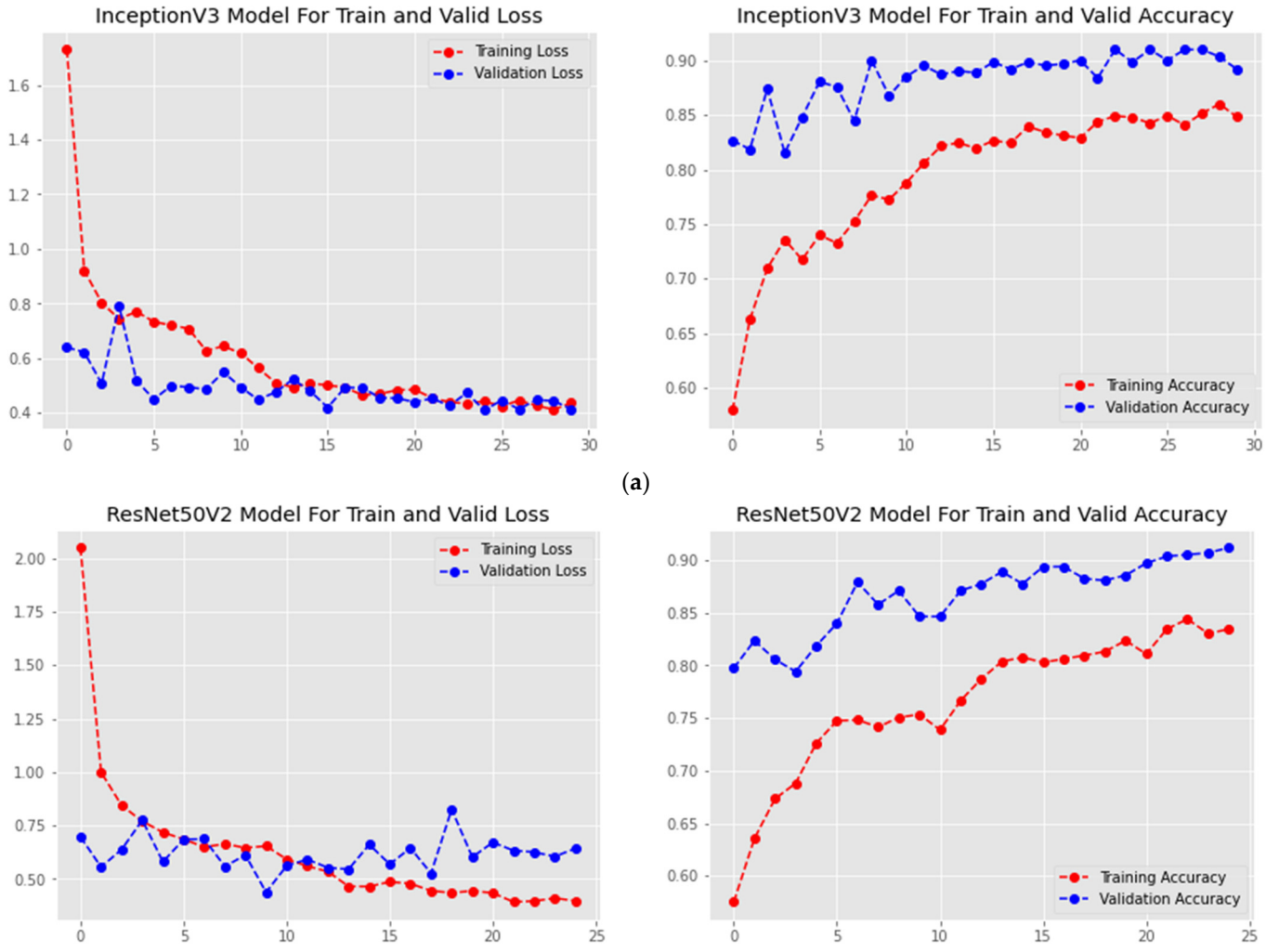


Figure 3. *Cont.*

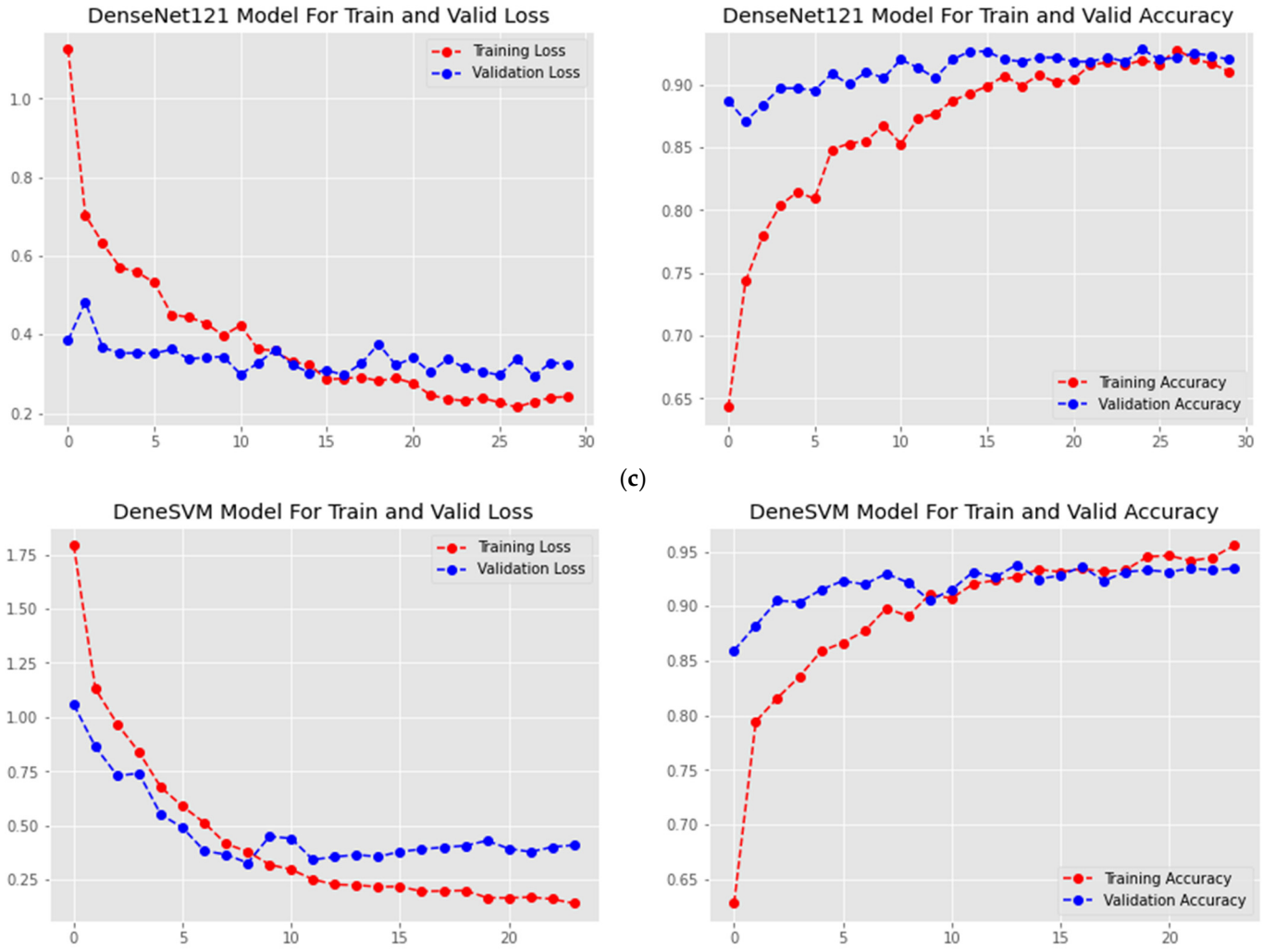
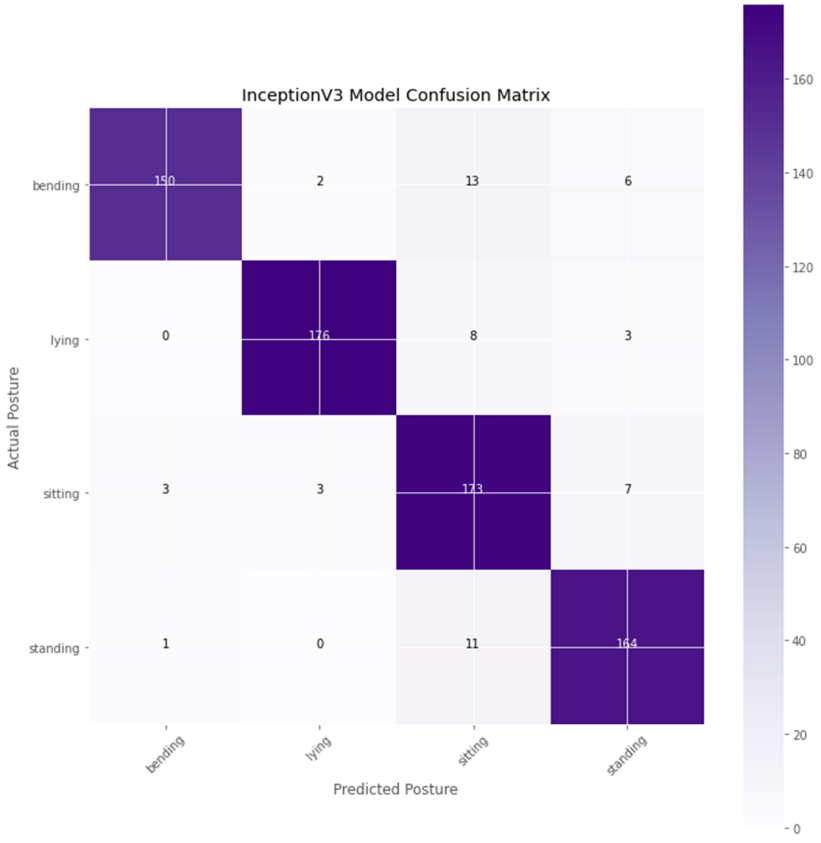
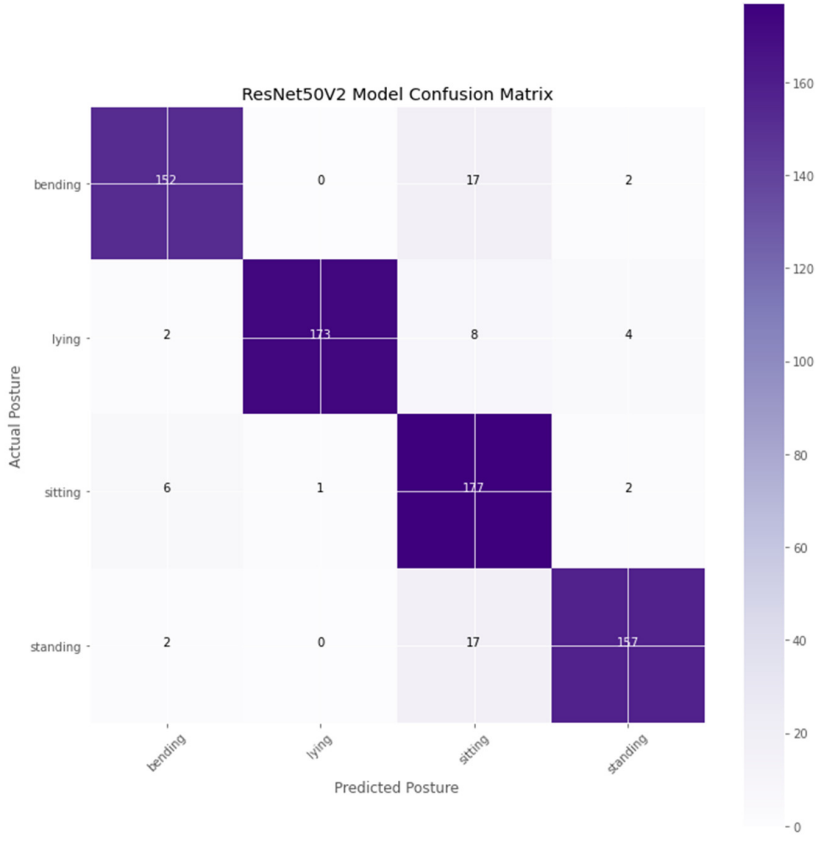


Figure 3. Accuracy and loss graphs of DTL models: (**a**) InceptionV3; (**b**) ResNet50; (**c**) DenseNet121; and (**d**) Proposed Inception-SVM.

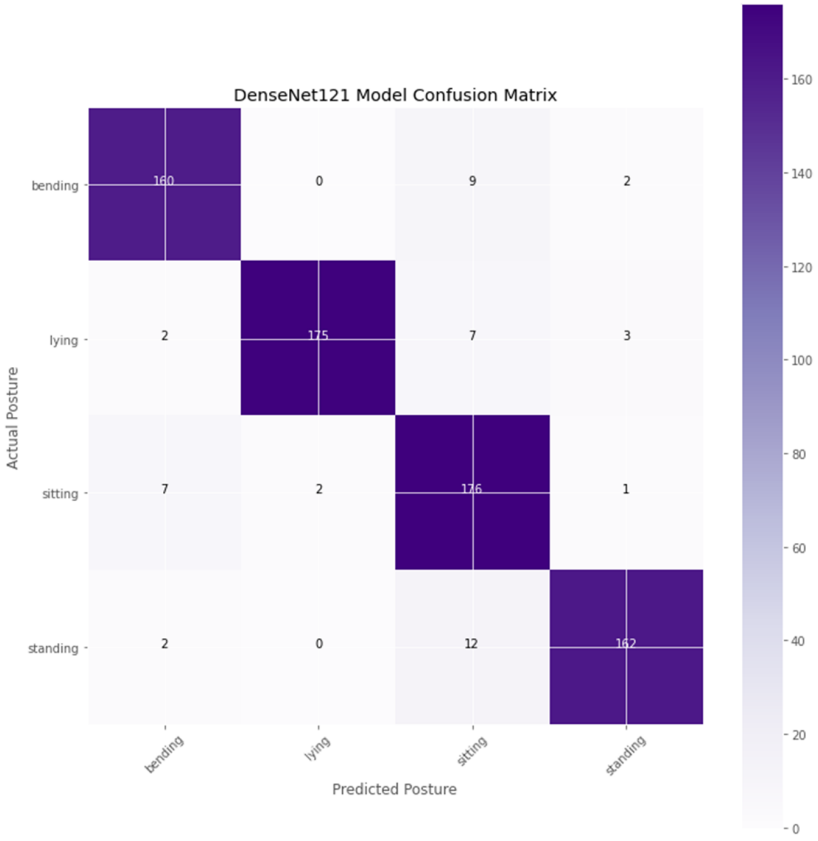


(**a**)

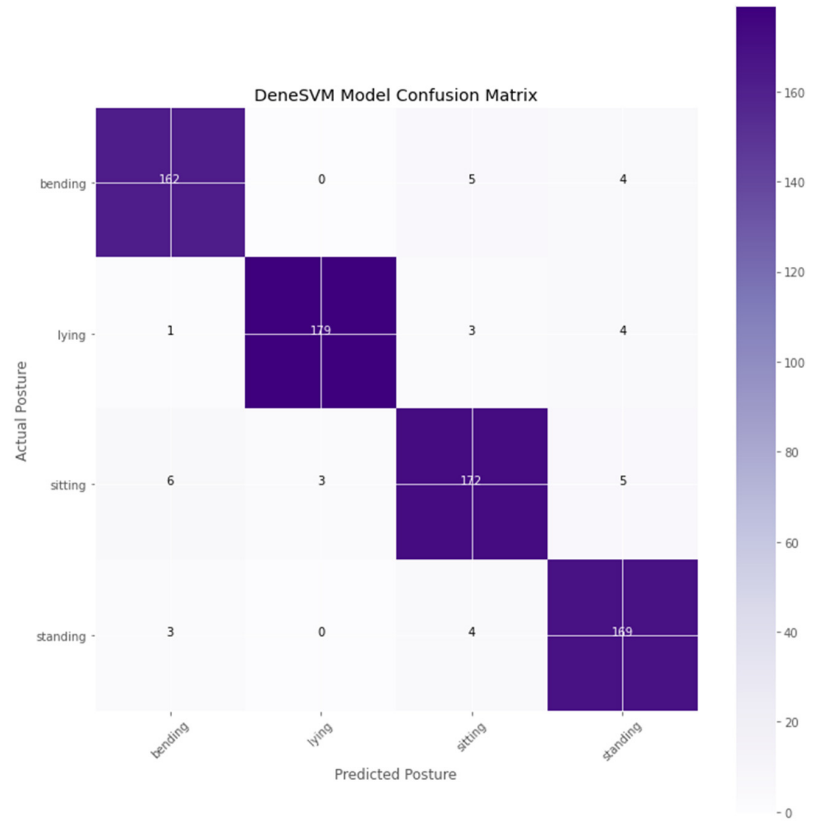


(**b**)

Figure 4. *Cont.*

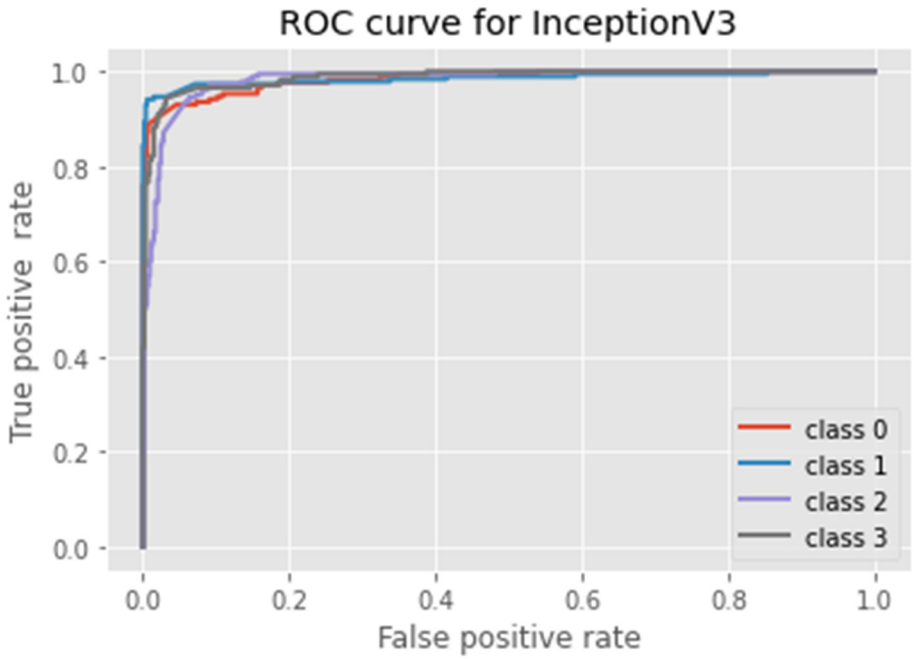


(**c**)

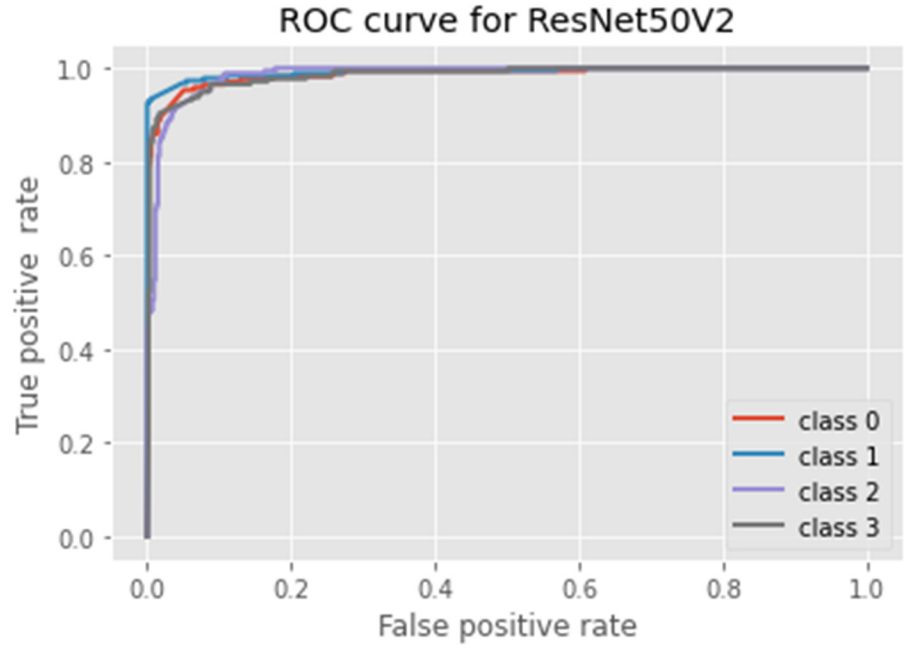


(**d**)

Figure 4. Confusion matrix of DTL models: (**a**) InceptionV3; (**b**) ResNet50V2; (**c**) DenseNet121; and (**d**) Proposed DeneSVM.

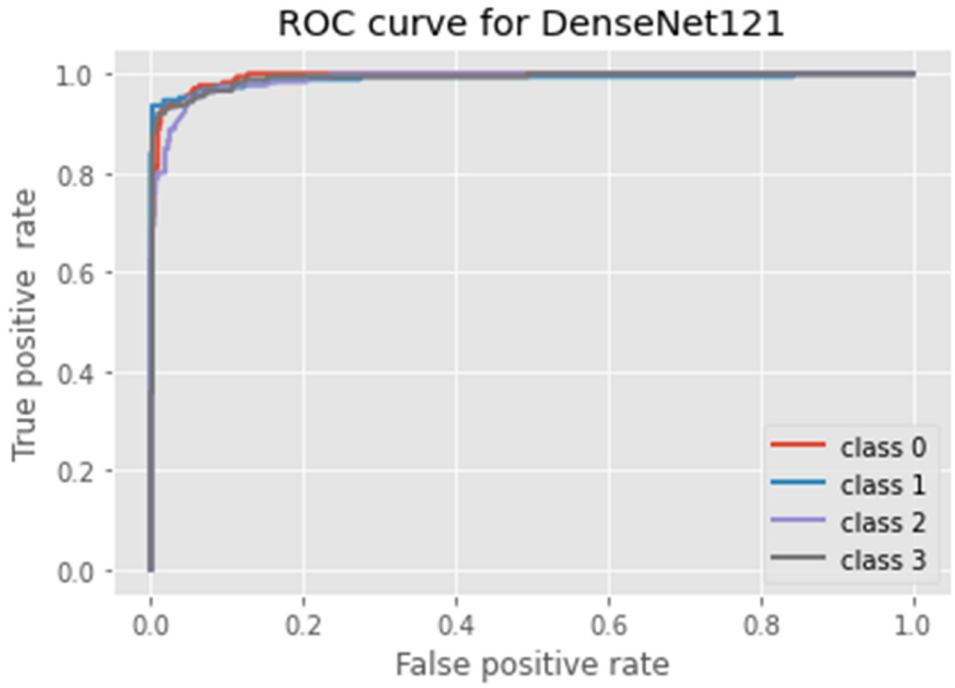


(**a**)

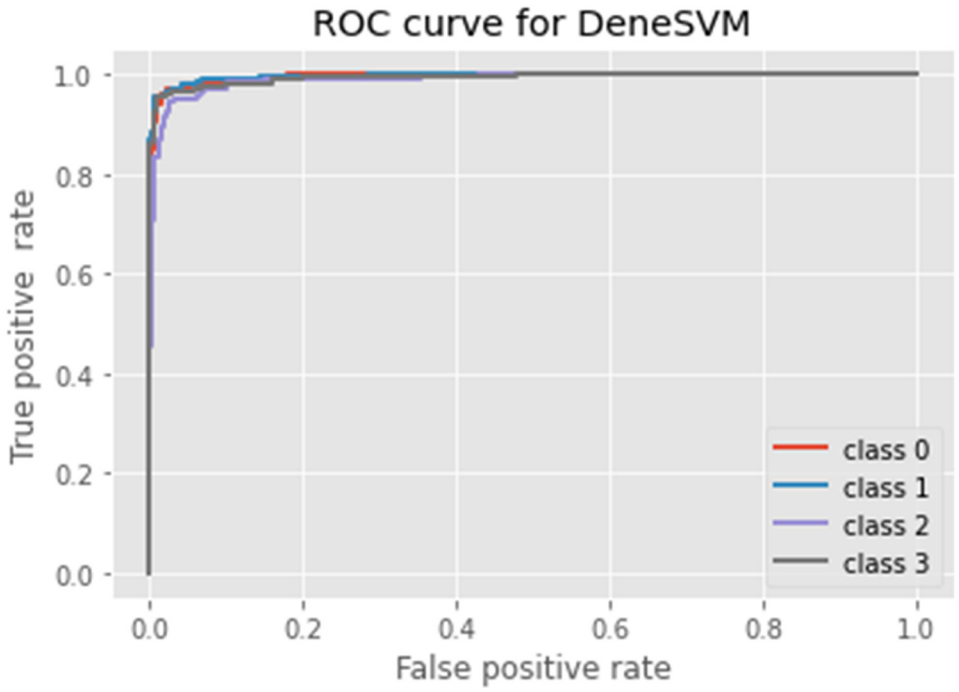


(**b**)

Figure 5. *Cont*.

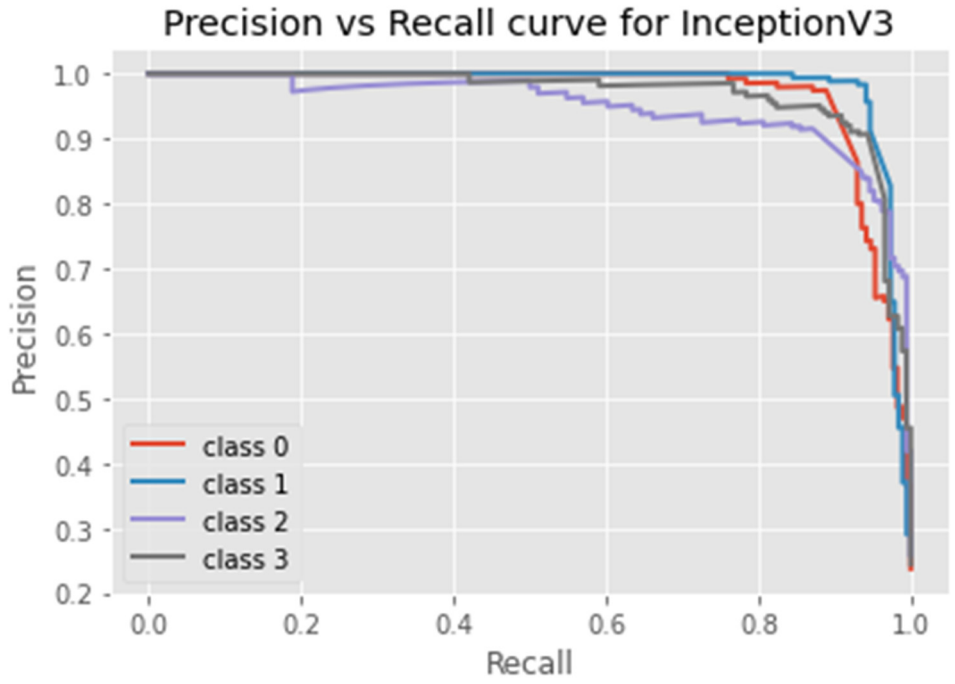


(**c**)

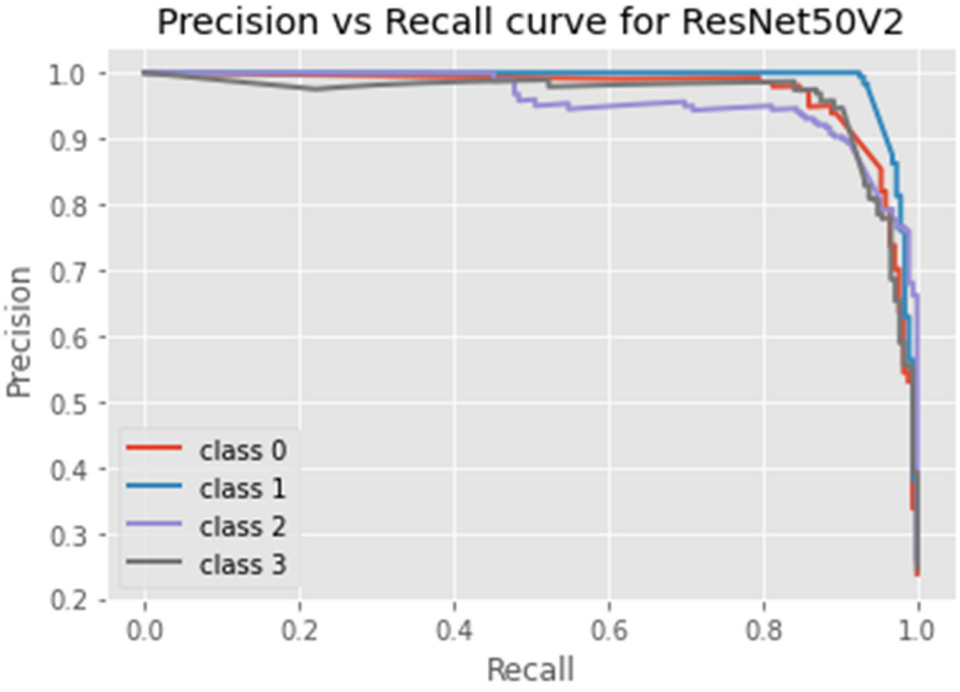


(**d**)

Figure 5. ROC curve of the DTL models: (**a**) InceptionV3; (**b**) ResNet50V2; (**c**) DenseNet121; and (**d**) proposed DeneSVM.

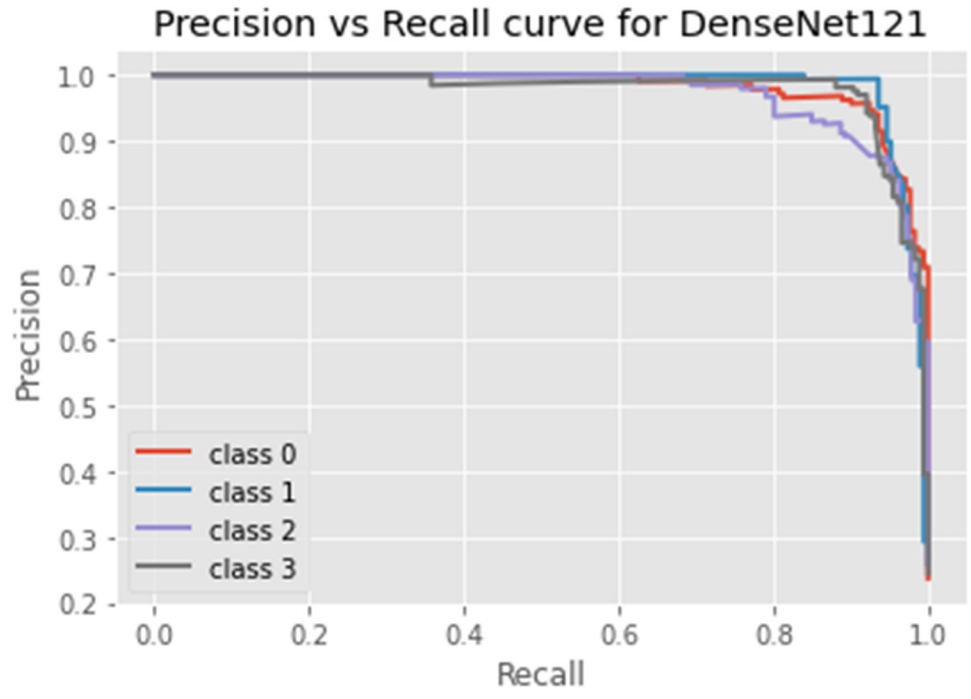


(**a**)

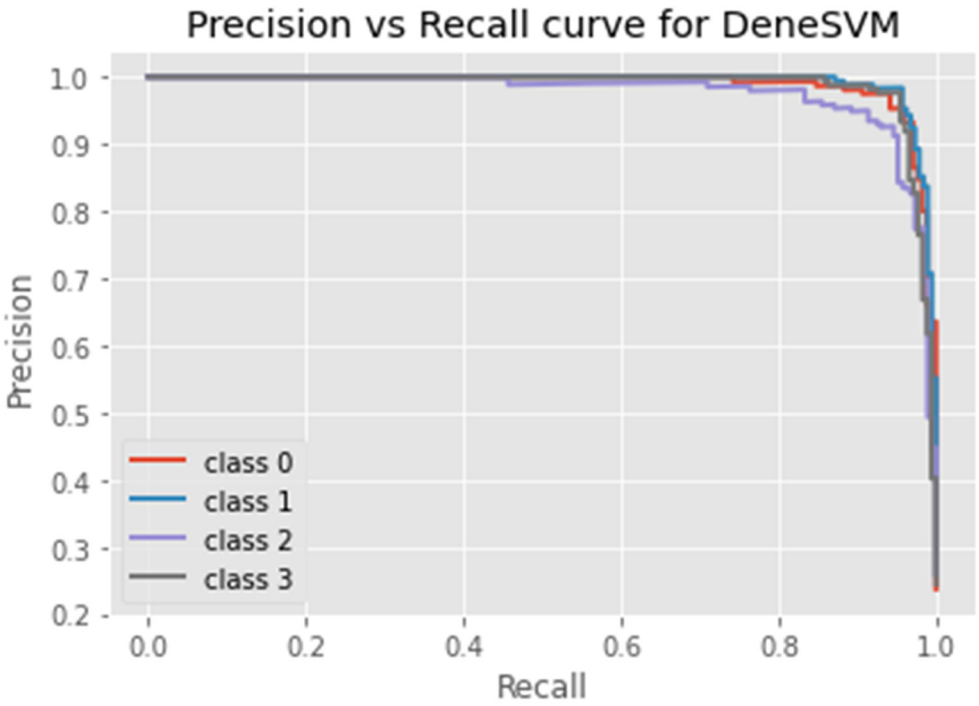


(**b**)

Figure 6. *Cont.*



(**c**)



(**d**)

Figure 6. Precision_Recall curve of DTL models: (**a**) InceptionV3; (**b**) ResNet50V2; (**c**) DenseNet121; and (**d**) proposed DeneSVM.

Table 6. Results of the performance evaluation obtained from Testing Dataset.

Model	Label	Precision	Recall	F1-Score	Accuracy	AUC
InceptionV3	Bending	0.97	0.88	0.92	0.92	98.46
	Lying	0.97	0.94	0.96		
	Sitting	0.84	0.93	0.88		
	Standing	0.91	0.93	0.92		
DenseNet-121	Bending	0.94	0.94	0.94	0.93	99.05
	Lying	0.99	0.94	0.96		
	Sitting	0.86	0.95	0.90		
	Standing	0.96	0.92	0.94		
ResNet-50V2	Bending	0.93	0.93	0.93	0.93	98.73
	Lying	1.00	0.93	0.96		
	Sitting	0.86	0.95	0.90		
	Standing	0.94	0.90	0.92		
DeneSVM (Proposed model)	Bending	0.96	0.94	0.95	**0.95**	**99.36**
	Lying	0.98	0.96	0.97		
	Sitting	0.90	0.96	0.93		
	Standing	0.96	0.93	0.94		

5. Discussion

The domain of ML for computer vision and image processing (IP) has been driven by DL networks. With the development of DL and IP, there is a chance to expand both research and application to include the detection and classification of human posture using position images. To effectively apply appropriate measures, rapid and precise models for human posture detection are necessary, thus reducing the risk of cardiovascular disease by addressing the problem of poor posture.

Deeper networks are much more reliable and easier to train, according to the latest DL research [23]. However, as the depth increases, more difficulties appear, including the problem of deterioration, disappearing contours, and internal covariate shift. The deep learning algorithm has also resulted in additional computing costs. For various designs, solutions to some of these issues have been put forth. These comprise skip connections [22], transfer learning [25], initialization strategies [32], optimization methods [33], and batch normalization [34].

Different image processing approaches [35], ML [36,37], and DL [26,38,39] have been deployed to the field of diagnosis using images. DL has discovered that DL is doing well. The use of the TL idea [26] shows that it increases accuracy while also speeding up execution.

The analysis of current DL techniques in the identification of human posture is carried out as a research expansion. A hybridized DeneSVM approach has been fine-tuned and InceptionV3, ResNet-50V2, and DenseNet-121 have been implemented and compared with the suggested approach. These designs have been effectively used in a variety of applications, including classification using ImageNet, Cifar 10, and Cifar 100.

The test results in Table 5 show that the suggested DeneSVM, InceptionV3, and DenseNet-121 outperformed ResNet-50V2 pretty well. This is evidence that deeper networks function more effectively than shallow networks. Figures 3–6 show how the four models implemented performed for the different human poses. Figure 6 indicates that the accuracy of the suggested approach increases as the number of epochs increases. At the same time, the loss decreases as the number of epochs increases. The accuracy of the model is 97.06% during training, 93.79% during validation, and 94.72% during testing. Consequently, the model achieved a training loss of 0.1318, the loss value on the validation data set is 0.4338, and the loss value on the test data set is 0.2918.

Figure 5 shows the ROC curve for all the implemented models, while the precision vs. recall curve for the four DTL models implemented is shown in Figure 6. It was deduced from the curves that the proposed DeneSVM model outperformed the baseline models with

an AUC score value of 99.36%, while among the baseline model, DenseNet121 performs better with an AUC score value of 99.05%. This shows that the hybridization of SVM with DenseNet121 improved the performance of the model with an increase of 0.31% AUC value over the baseline DenseNet121 model, 0.63% over the ResNet50V2 model, and 0.9% over InceptionV3. Similarly, the proposed DeneSVM is better with an increase of 2% of test data accuracy over DenseNet121 and 3% over InceptionV3 and ResNet50V2.

Compared to the other architectural designs, DeneSVM (DenseNet-121 with SVM) is faster to train. ResNet also performed well, although its training duration is much longer than that of DeneSVM, InceptionV3 and DenseNet. Incredibly deep networks are potentially more efficient and require fewer weights, as shown by DenseNet, InceptionV3, and ResNet architectures.

6. Conclusions

In this study, the existing DTL for image-based human posture detection is fine-tuned, hybridized, and evaluated. Hybridized DeneSVM, InceptionV3, DenseNet-121, and ResNet-50V2 are among the models investigated. According to the implementations, the DeneSVM model tends to provide a strong increase in accuracy with an increasing number of epochs, without showing any signs of performance problems or overfitting. The DeneSVM approach also performs well in classification and detection exhibitions, with reasonable computation costs and a significantly smaller number of parameters. DeneSVM achieves a test accuracy score of 94.72% for the 30th epoch, and a test accuracy of 95% on the testing dataset, beating the remaining approaches implemented. DeneSVM is therefore a good approach to classifying human posture.

Although the design performs well, further study is needed to reduce the computing time. In addition, different DTL models may be used to categorize human postural datasets. In the future, the proposed DeneSVM model will be applied to realistic images or video streams to see how the model works in the surrounding environment. The model is also proposed to be applied to different human posture datasets such as COCO and MPII.

Author Contributions: Conceptualization, R.M.; methodology, R.O.O.; software, R.O.O.; validation, R.O.O., R.M. and R.D.; formal analysis, R.O.O., R.M., R.D. and S.M.; investigation, R.O.O., R.M. and R.D.; resources, R.M.; data curation, R.O.O.; writing—original draft preparation, R.O.O., and R.M.; writing—review and editing, S.M. and R.D.; visualization, R.O.O.; supervision, R.M.; project administration, S.M. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive external funding.

Data Availability Statement: The data presented in this study are openly available in the Kaggle repository https://ieee-dataport.org/accessed on 4 August 2022. https://www.kaggle.com/datasets/deepshah16/silhouettes-of-human-posture accessed on 4 August 2022. The codes required to execute this study have already been posted to the GitHub repository and can be found in the repository: https://github.com/Roseybaby/DeneSVM-on-Silhouettes-Data-for-HPD accessed on 4 August 2022.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Verma, A.; Suman, A.; Biradar, V.G.; Brunda, S. Human Activity Classification Using a Deep Convolutional Neural Network. In *Recent Advances in Artificial Intelligence and Data Engineering*; Springer: Singapore, 2022; pp. 41–50.
2. Ogundokun, R.O.; Maskeliunas, R.; Misra, S.; Damaševičius, R. Improved CNN Based on Batch Normalization and Adam Optimizer. In *International Conference on Computational Science and Its Applications*; Springer: Cham, Switzerland, 2022; pp. 593–604.
3. Le, N.Q.K.; Ho, Q.T. Deep transformers and a convolutional neural network to identify DNA N6-methyladenine sites in genomes of cross-species. *Methods* **2022**, *204*, 199–206. [CrossRef] [PubMed]
4. Danilatou, V.; Nikolakakis, S.; Antonakaki, D.; Tzagkarakis, C.; Mavroidis, D.; Kostoulas, T.; Ioannidis, S. Outcome Prediction in Critically Ill Patients with Venous Thromboembolism and/or Cancer Using Machine Learning Algorithms: External Validation and Comparison with Scoring Systems. *Int. J. Mol. Sci.* **2022**, *23*, 7132. [CrossRef] [PubMed]

5. Ogundokun, R.O.; Maskeliūnas, R.; Damaševičius, R. Human Posture Detection Using Image Augmentation and Hyperparameter-Optimized Transfer Learning Algorithms. *Appl. Sci.* **2022**, *12*, 10156. [CrossRef]
6. Le, N.Q.K. Potential of deep representative learning features to interpret sequence information in proteomics. *Proteomics* **2021**, *22*, e2100232. [CrossRef] [PubMed]
7. Ali, F.; El-Sappagh, S.; Islam, S.R.; Kwak, D.; Ali, A.; Imran, M.; Kwak, K.S. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf. Fusion* **2020**, *63*, 208–222. [CrossRef]
8. Choi, Y.A.; Park, S.J.; Jun, J.A.; Pyo, C.S.C.; Cho, K.H.; Lee, H.S.; Yu, J.H. Deep learning-based stroke disease prediction system using real-time biosignals. *Sensors* **2021**, *21*, 4269. [CrossRef] [PubMed]
9. Pan, Y.; Fu, M.; Cheng, B.; Tao, X.; Guo, J. Enhanced deep learning-assisted convolutional neural network for heart disease prediction on the platform of medical things platform. *IEEE Access* **2020**, *8*, 189503–189512. [CrossRef]
10. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multiperson 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.
11. Mehr, H.D.; Polat, H. Recognition of human activity in the smart home with the deep learning approach. In Proceedings of the 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), Istanbul, Turkey, 25–26 April 2019; pp. 149–153.
12. Du, H.; He, Y.; Jin, T. Transfer learning for human activities classification using micro-Doppler spectrograms. In Proceedings of the 2018 IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, China, 26–28 March 2018; pp. 1–3.
13. Shi, X.; Li, Y.; Zhou, F.; Liu, L. Human activity recognition is based on the deep learning method. In Proceedings of the 2018 International Conference on Radar (RADAR), Brisbane, Australia, 27–31 August 2018; pp. 1–5.
14. Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Human activity detection from RGBD images. In Proceedings of the Workshops at the 25th AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
15. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732.
16. Laptev, I.; Marszalek, M.; Schmid, C.; Rozenfeld, B. Learning realistic human actions from movies. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 24–26 June 2018; pp. 1–8.
17. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning is applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
18. Abhishek Kumar Indian Institute of Information Technology Kottayam. EbinDeni Raj Indian Institute of Information Technology Kottayam. Available online: https://ieee-dataport.org/ (accessed on 5 September 2022).
19. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 630–645.
23. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
24. Keleş, S.; Günlü, A.; Ercanli, I. Estimation of the carbon of the aboveground stand by combining Sentinel-1 and Sentinel-2 satellite data: A case study from Turkey. In *Forest Resources Resilience and Conflicts*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 117–126.
25. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
26. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]
27. Ho, E.S.L.; Chan, J.C.P.; Chan, D.C.K.; Shum, H.P.H.; Cheung, Y.; Yuen, P.C. Improving posture classification accuracy for depth sensor-based human activity monitoring in smart environments. *Comput. Vis. Image Underst.* **2016**, *148*, 97–110. [CrossRef]
28. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
29. Zhang, C.; Ma, Y. (Eds.) *Ensemble Machine Learning: Methods and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
30. Chu, G. Machine learning for automation of Chromosome based Genetic Diagnostics. *Digit. Vetensk. Ark.* **2020**, *832*, 46.
31. Glorot, X.; Bordes, A.; Bengio, Y. Deep-sparse rectifier neural networks. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323, JMLR Workshop and Conference Proceedings.
32. Mishkin, D.; Matas, J. All you need is a good idea. *arXiv* **2015**, arXiv:1511.06422.
33. Le, Q.V.; Ngiam, J.; Coates, A.; Lahiri, A.; Prochnow, B.; Ng, A.Y. On optimization methods for deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, WA, USA, 28 June 2011; pp. 265–272.

34. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
35. Samanta, D.; Chaudhury, P.P.; Ghosh, A. Scab disease detection of potatoes using image processing. *Int. J. Comput. Trends Technol.* **2012**, *3*, 109–113.
36. Athanikar, G.; Badar, P. Potato leaf disease detection and classification system. *Int. J. Comput. Sci. Mob. Comput.* **2016**, *5*, 76–88.
37. Wang, H.; Li, G.; Ma, Z.; Li, X. Application of neural networks to the recognition of image of plant diseases. In Proceedings of the 2012 International Conference on Systems and Informatics (ICSAI2012), Shandong, China, 19–20 May 2012; pp. 2159–2164.
38. Ogundokun, R.O.; Misra, S.; Douglas, M.; Damaševičius, R.; Maskeliūnas, R. Medical Internet-of-Things Based Breast Cancer Diagnosis Using Hyperparameter-Optimized Neural Networks. *Future Internet* **2022**, *14*, 153. [CrossRef]
39. Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep neural network-based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* **2016**, *2016*, 3289801. [CrossRef] [PubMed]