Handwriting Recognition in Down Syndrome Learners Using Deep Learning Methods

Kirsty-Lee Walker[a] and Tevin Moodley[b]
University of Johannesburg, Kingsway Avenue and University Rd, Auckland Park, Johannesburg 2092, South Africa

Keywords: Deep Learning, VGG16, InceptionV2, Xception, Down Syndrome, Handwriting Recognition.

Abstract: The Handwriting task is essential for any learner to develop as it can be seen as the gateway to further academic progression. The classification of Handwriting in learners with down syndrome is a relatively unexplored research area that has relied on manual techniques to monitor handwriting development. According to earlier studies, there is a gap in how down syndrome learners receive feedback on handwriting assignments, which hinders their academic progression. This research paper employs three deep learning architectures, VGG16, InceptionV2, and Xception, as end-to-end methods to categorise Handwriting as down syndrome or non-down syndrome. The InceptionV2 architecture correctly identifies an image with a model accuracy score of 99.62%. The results illustrate the manner in which the InceptionV2 architecture is able to classify Handwriting from learners with down syndrome accurately. This research paper advances the knowledge of which features differentiate a down syndrome learner's Handwriting from a non-down syndrome learner's Handwriting.

1 INTRODUCTION

Deep learning is considered to have many applications in a wide range of industries, including hospitality, agriculture, energy, and many more (Almalaq and Edwards, 2017). Breaking down raw information into numerous layers of pre-processed input and then extracting higher-level features is a ground-breaking innovation (Priatama et al., 2022). Deep learning has also made advances in the education sector, such as Handwriting, which requires much effort and consistent feedback to master. Writing by hand boosts self-esteem, promotes better memory recall, and helps with reading and speaking (Engel-Yeger et al., 2009). According to prior studies, down syndrome learners have trouble writing as they can have short, stubby fingers and low muscle tone, which impacts their gross and fine motor skills (Chumlea et al., 1979).

With the help of an end-to-end deep learning model, this research paper aims to distinguish handwritten images as down syndrome or non-down syndrome. Three deep learning architectures are employed, VGG16, InceptionV2, and Xception, to expand the field of study and demonstrate how deep learning can be used to solve practical issues. This research paper explores the proposed architectures to determine which architecture is best suited to the domain. The problem background is discussed in Section 2, which covers related studies. Section 3 describes the attributes and the manner in which the dataset was constructed. The different architectures proposed for this research are unpacked in section 4. Finally, sections 5 and 6 breaks down the findings along with the performance and difficulties that the various architectures face.

[a] https://orcid.org/0000-0001-8342-9731
[b] https://orcid.org/0000-0002-5330-3908

2 PROBLEM BACKGROUND

Deep learning has been around for a while (Schmidhuber, 2015), but there are not many applications in down syndrome education especially when it comes to Handwriting. According to earlier studies, the classification of Handwriting for learners with down syndrome has depended on labour-intensive, error-prone manual methods of data collection and classification.

2.1 Related Works

Tsao et al. proposed Ecriture Suite software and R software to evaluate graphomotor control, which relates to the muscular movement required to perform

Walker, K. and Moodley, T.
Handwriting Recognition in Down Syndrome Learners Using Deep Learning Methods.
DOI: 10.5220/0011888500003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 936-943
ISBN: 978-989-758-634-7; ISSN: 2184-4321

writing tasks (Tsao et al., 2017). The graphomotor control evaluation was based on six spatiotemporal indices; handwriting speed, number of pauses, pause duration, stroke duration, stroke length, and pen pressure. In the study, 72 people participated, where 24 were down syndrome, 24 were of the same development age, and the remaining participants were of the same chronological age (Llamas et al., 2017). Chronological age is the age as determined by the person's date of birth. Developmental age describes the degree to which an individual's physical and mental development corresponds with typical developmental milestones (Eaton et al., 2014). The participants wrote letters on a tablet, and the handwritten images were analysed using Ecriture Suite software (Tsao et al., 2017).

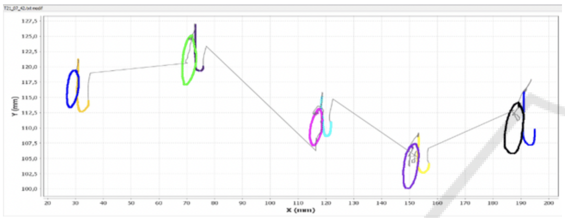

Figure 1: A figure depicting the Ecriture Suite software. The graph illustrates the participant's stroke or pen lift, the colour code is used to indicate the different strokes (in colour) and the pen lifts between these strokes (in grey) (Tsao et al., 2017).

The graph in figure 1 illustrated pen pressure and handwriting speed over time (mm/s). R software was used to calculate each indicator's individual coefficient of variation (ICV). The most significant result was that the standard deviation for handwriting speed for a down syndrome learner was 38.6, whereas the development age was 62.79. Similarly, the number of pauses for a down syndrome learner was 3.15, and the developmental age was 0.52. Researchers concluded that there is a developmental delay in writing acquisition for learners and adults with down syndrome. The developmental delay confirms that an autonomous way to identify handwritten images must be identified.

Patton & Hutton proposed the Handwriting Without Tears (HWT) software to encourage down syndrome learners to write (Patton and Hutton, 2017). In the study, 46 down syndrome learners participated in the HWT handwriting curriculum (Patton and Hutton, 2017). There were 7 HWT group meetings during the eight-month curriculum. After the eight months were concluded, the participants subsequently answered a questionnaire on their experiences with the HWT handwriting curriculum.

Based on the findings, an observation tool was created to assess a learner's engagement, interest, capacity to stay on task, and fine motor coordination. According to the findings, down syndrome learners would be more likely to participate in handwriting promotion activities when learning through hands-on, multi-sensory techniques (Patton and Hutton, 2017). The researchers concluded that techniques for evaluating Handwriting in learners with down syndrome need to be more reliable and robust (Patton and Hutton, 2017). This study underlines the need for down syndrome learners to access a handwriting recognition system that can provide feedback and promote long-term handwriting learning.

More recent works aim to recognise characters in handwritten images using a multilayer perceptron neural network (Adamu et al., 2017). To reduce noise, pre-processing techniques such as grey scaling, noise reduction, binarisation, skeletonisation, normalisation, and segmentation were applied (Adamu et al., 2017). During the feature extraction phase, the features were mapped to a feature vector and then classified as an individual input character (the alphabet or special letters) (Adamu et al., 2017). The model accurately identified all 26 English alphabet letters with an accuracy rate of 95.0%. The study highlighted that employing a multilayer perception neural network decreases the time and expense associated with training to recognise handwritten characters (Adamu et al., 2017). Deep learning models can be used to accurately classify handwriting (Adamu et al., 2017). This research paper compares the VGG16, InceptionV2, and Xception to determine which architecture best suits the problem domain. Each architecture will be fined tuned on the handwriting dataset.

3 EXPERIMENT SETUP

The English alphabet is well known and simplistic. The capital letters O, A, T, X, C, and F are the simplest to write, whereas D, G, J, Y, and Z are the most challenging (Puranik et al., 2013). Figure 2 shows how a learner with down syndrome struggles with all the letters, demonstrating how challenging Handwriting is for a down syndrome learners. Therefore a model that can categorise whether Handwriting derives from a down syndrome or non-syndrome learner may help in giving prompt feedback on areas that need improvement.

The non-down syndrome handwriting data is collected from GNHK (Good Note Handwriting Kollection) repository (Lee et al., 2021), which contains images of different ages, handwriting neatness and types (reports, shopping lists, worksheets, study notes, di-

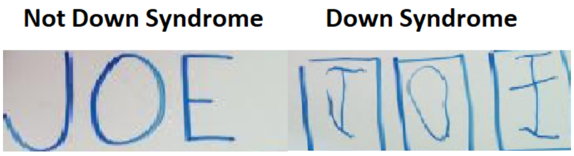

Figure 2: An image of a down syndrome learner's Handwriting (on the right) and non-down syndrome learner's Handwriting (on the left) (Association, 2017).

agrams, and letters). As a result of there being no dataset available for down syndrome a manual approached was used to collect the down syndrome images. A software tool called ParseHub was used to scraped images and most of the images sourced came from an online community called Pals. Other images contain text relating to letters and worksheets. This dataset was used to research various ways of efficiently detecting images such as down syndrome or non-down syndrome using deep learning architectures. The dataset was manually inspected to remove images that did not meet the specifications or if the images could not be differentiated comprehensively. The resulting number of images in the dataset was 200 unique images. It was noted that there is a lack of images relating to the Handwriting of down syndrome learners, which contributed to the small dataset size used in this study. The dataset is divided using an 80/20 split, where 160 images were used for training and 40 images for testing. Through testing and validation, the image size chosen was 128x128. The dataset used in this study can be accessed using the following link DownSyndrome.

4 METHODS

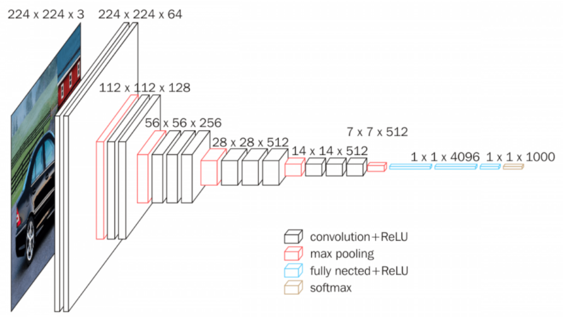

Figure 3: A figure representing the VGG16 architecture (Anwar, 2019).

VGG16 is an improvement on the AlexNet architecture whereby the VGG16 architecture replaces large kernel filters with multiple 3x3 kernel-sized filters (Anwar, 2019). Karen Simonyan and Andrew Zisserman introduced the VGG16 architecture and subsequently won the 2014 ImageNet object recognition challenge (Yu et al., 2016). The VGG16 architecture has seen much success due to its number of layers, which is 16 layers deep. Figure 3, depicts the structure of the VGG16 architecture. The VGG16 architecture takes in an RGB image which passes through the first stack of two convolution layers with a receptive size of 3x3, followed by ReLu (rectified linear activation function) (Yu et al., 2016). To prevent the negative values from being sent to the subsequent layers, ReLu is then applied. Each of these two layers contains 64 filters, each with a convolution stride and padding of 1 pixel to maintain the spatial resolution (Anwar, 2019). The activation map runs through spatial max pooling with a 2x2 pixel window and a 2-pixel stride (Yu et al., 2016). The size of the first stack is (112x112x64), and the second stack has a size of (56x56x128). Following the third stack, there are three convolutional layers and a max pooling layer resulting in a size of (28x28x512). Two stacks of three convolution layers with 512 filters are added after the third stack. The output of both of these stacks is (7x7x512). Three completely connected layers, separated by a flattening layer, are added after the sixth stack. The last has an output layer with 1000 neurons, while the first two contain 4096 neurons each. A softmax activation layer for categorical classification is placed after the output layer. The biggest flaw of the VGG16 architecture is that it is slower than newer architectures (Anwar, 2019).

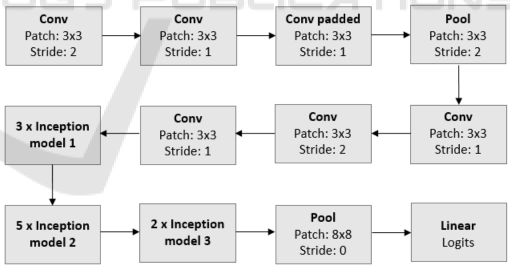

Figure 4: A figure representing the InceptionV2 architecture (Nguyen et al., 2018).

InceptionV2 is an enhanced version of InceptionV1 architecture introduced in 2014 by Szegedy et al. (Szegedy et al., 2015). InceptionV2 utilises factorisation in the convolution layer to address the overfitting problem. To benefit from batch normalisation, InceptionV2 removes local response normalisation and uses auxiliary classifiers as regularisers (Szegedy et al., 2015). In a conventional convolutional neural network, the output from the previous layer serves as the input for the following layer, and so on until the prediction (Szegedy et al., 2015). In figure 4 the InceptionV2 design is 48 layers deep. In some circumstances, it has been observed that the deeper the architecture, the closer it can come to the ideal function (Brownlee, 2019). Similarly, a deeper model performs better due to the ability to learn a more complicated, non-linear function (Brownlee, 2019). Each of the 11 modules of the InceptionV2 architectures consists of pooling layers and convolutional filters with ReLu serving as the activation function (Pandit et al., 2020). Two 3x3 convolutions in InceptionV2 replace the 5x5 convolutions. The convolution filter reduces the original input before applying various size filters (1x1, 3x3, and 5x5) and a max pooling layer (Szegedy et al., 2015). The inception block sends the input from the preceding layers to four separate operations concurrently. Each layer has a higher accuracy than the layers before it due to the concatenation of the outputs before they are delivered to the next layer (Agarwal, 2017).

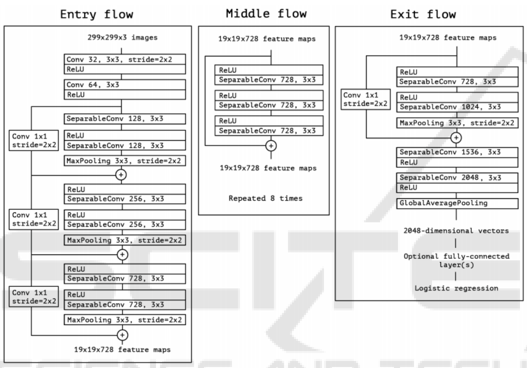

Figure 5: A figure representing the Xception architecture (Akhtar, 2021).

Extreme Inception, or Xception, is a network first introduced by Francois Chollet (Chollet, 2017). According to Akhtar et al., Xception is an addition to the Inception architecture that uses depthwise separable convolutions in place of the normal Inception modules (Akhtar, 2021). These depthwise separable convolutions aid in reasonably accurately classifying millions of images (Muhammad et al., 2021). In figure 5, the Xception architecture is seen, where the data passes through the entering flow, then the middle flow, repeated eight times, and finally passes through the exit flow (Akhtar, 2021). Each depth map is first subjected to the filters, which are then applied across the depth to compress the original space using 1x1 convolutions (Muhammad et al., 2021). One notable distinction between the Inception model and Xception is that although Xception does not introduce any non-linearity, the Inception model is followed by ReLu non-linearity (Muhammad et al., 2021).

The smaller dataset in this research paper makes transfer learning suitable. Transfer learning reuses a previously learned model to solve a new problem which shortens training time and conserves resources by eliminating the need to train numerous models from scratch to carry out related tasks (Shu, 2019). Although transfer learning has several advantages, including small datasets, speed, and computational complexity, there are drawbacks as well (Shu, 2019). Transfer learning performance diminishes due to its inability to stop the negative transfer (Shu, 2019). Similarly, transfer learning can lead to overfitting when a new model acquires traits from training data that degrade its performance. Overfitting occurs when the training data fits exactly against the training data (Ying, 2019). Since overfitting was not encountered in this research, there is no need to perform regularisation, decrease network capacity or add dropout layers.

Table 1: A table illustrating the top-1 and top-5 accuracy for the VGG16, InceptionV2, and Xception architectures on the ImageNet dataset to justify the selection of architectures in the study (Robert and Thomas.,).

Model	**Top-1**	**Top-5**
VGG16	74.40%	91.90%
InceptionV2	74.80%	92.00%
Xception	79.00%	94.50%

Table 1 examines the top-1 and top-5 accuracy of each architecture proposed in this research to support the architecture of choice. The ImageNet dataset is used to pre-train and benchmark the VGG16, InceptionV2, and Xception architectures to demonstrate each architecture's performance on a generalised dataset. With a top-1 accuracy of 79.0% and top-5 accuracy of 94.50%, the Xception architecture provides the best performance when used with ImageNet, as seen in table 1. To determine whether Xception performs the best in this research domain results from table 1 will be further examined in the results and discussion section. Additionally, to ensure the architectures are compared fairly, a global average pooling layer and a dense output layer with softmax activation is added to predict a multinominal probability distribution (Wani et al., 2020). The adam optimiser is then used to assist the architectures in classifying noise data, and sparse gradients (Wani et al., 2020). Each architecture shuffle is set to false, the weights are set to none, and the input tensor is set to none. No other parameters are altered. Data augmentation techniques are applied to each architecture. Through testing and validation, A shear zoom of 0.2, and image rescaling of 1./255 is applied (Wani et al., 2020). The shear zoom range slants the images to a 20-degree angle, and rescaling of 1./22 ensures the images are not distorted (Wani et al., 2020).

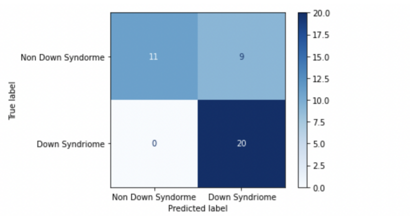

Figure 6: A figure illustrating the VGG16 architecture confusion matrix highlighting 9 images incorrectly predicted as down syndrome.

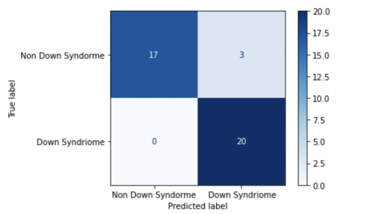

Figure 7: A figure illustrating the InceptionV2 architecture confusion matrix highlighting 3 images incorrectly predicted as down syndrome.

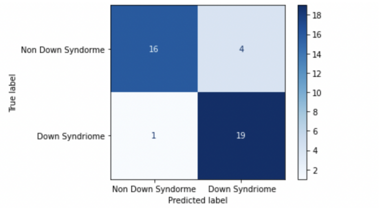

Figure 8: A figure illustrating the Xception architecture confusion matrix where 4 images are incorrectly predicted as down syndrome.

5 RESULTS

To ensure a fair and comparable baseline, each architecture was run five times in table 2. The IcnceptionV2 architecture performs the best based on training data, with an accuracy of 99.56% and a loss of 0.01265. The Xception architecture achieved an accuracy of 96.76% and a loss of 0.07340. The VGG16 has a loss of 0.41000 and an accuracy of 87.40%.

Table 2: A table comparing accuracy and loss for the VGG16, InceptionV2, and Xception architectures.

Model	Accuracy	Loss
VGG16	87.40%	0.41000
InceptionV2	99.56%	0.01265
Xception	96.76%	0.07340

The results in table 3 reveal that the InceptionV2 is the best-suited architecture for the problem domain with the highest f1 score, precision and recall. The confusion matrices in figure 6, figure 7, figure 8 are obtained after five runs, the average is computed for each metric.

Table 3: A table comparing the precision, recall, f1-score for the VGG16, InceptionV2, and Xception architectures.

Model	Precision	Recall	F1 Score
VGG16	89.00%	87.20%	88.10%
InceptionV2	97.40%	96.80%	97.10%
Xception	91.00%	89.40%	90.20%

In figure 6, out of 40 test samples, 11 images are predicted correctly as non-down syndrome and 20 images are correctly predicted as down syndrome. Similarly, 9 images are incorrectly predicted as down syndrome.

Based on figure 7, 17 images were predicted correctly as non-down syndrome and 20 images were correctly predicted as down syndrome. Similarly, 3 images were incorrectly predicted as being down syndrome Handwriting.

In Figure 8, 16 images are predicted correctly as non-down syndrome, and 19 images are correctly predicted as down syndrome. Similarly, 4 images are incorrectly predicted as down syndrome. Similarly, 1 image is incorrectly predicted as non-down syndrome. In all three confusion matrices, figure 9 show images that were the most commonly misclassified images.

In Figure 10 the InceptionV2 network achieved the highest probability of 99.62% that the same unseen image is an example of the down syndrome class, which may be a result of the high f1-score of 97.10%. Simiarly, The Xception network had the probability of 99.62% and the VGG16 has the probability of 86.64% for the same unseen image.

6 DISCUSSION OF RESULTS

This research paper aims to recognise whether handwritten images are from a down syndrome or non-down syndrome learner. Table 2 shows that the InceptionV2 architecture outperforms the VGG16 and Xception architecture in terms of performance and ac-

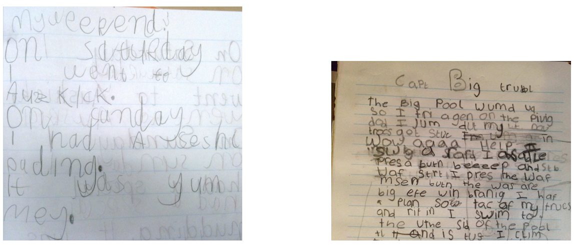

Figure 9: Samples of the most misclassified images.

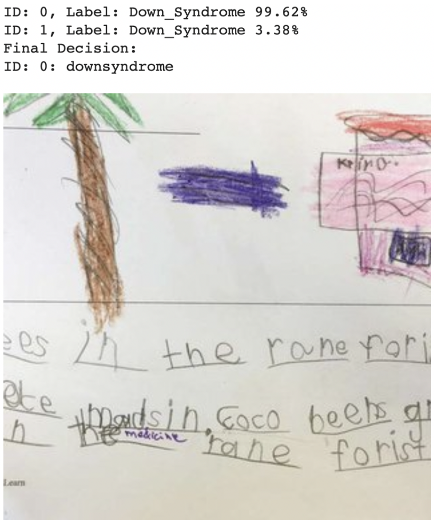

Figure 10: An image depicting the InceptionV2 confidence of classifying an unseen image.

curacy after each architecture is run five times to get the mean. The precision, recall, and f1-scores of the InceptionV2 and the Xception network are comparable in table 3. The VGG16 architecture performs the worst which is supported using the top-1 and top-5 accuracy as seen in table 1. Although the VGG16 model can be more discriminative due to its smaller window size (Theckedath and Sedamkar, 2020) the VGG16 architecture suffers from the vanishing gradient issue. The vanishing gradient happens when the gradient is significantly smaller, preventing the weight from changing its value (Brownlee, 2019). The vanishing gradient problem makes it challenging to relearn and fine-tune parameters early on which stops the neural network from learning completely. In this research paper steps are taken to use reLu to mitigate the vanishing gradient problem.

The Xception architecture achieves the highest top-1 and top-5 accuracy in table 1. However, the Xception architecture did not have the best accuracy for the specified research domain. It has been noted that the Xception design works better on larger datasets, which may account for its lower performance within this study. One other possible reason for its poorer performance is that it has more depthwise separable convolution on kernels of various sizes (Muhammad et al., 2021). The results in table 2 illustrate the minor performance difference between the InceptionV2 and Xception architecture. Therefore, it is challenging to pinpoint the exact reason for the InceptionV2 architecture yielding better performance.

One plausible reason for the InceptionV2 performance is that it has fewer parameters. Fewer parameters make it better for classifying smaller datasets (Agarwal, 2017). The InceptionV2 architecture also offers a way to overcome overfitting. The InceptionV2 architecture uses factorisation in the convolution layer to reduce the overfitting problem (Szegedy et al., 2015). Similarly, the InceptionV2 architecture uses auxiliary classifiers as regularises, improving the convergence during training by mitigating the vanishing gradient problem (Szegedy et al., 2015). The configuration of the InceptionV2 architecture consists of 1x1 filters followed by convolutional layers with different filter sizes applied simultaneously. This makes the InceptionV2 architecture ideal for classifying down syndrome, and non-down syndrome Handwriting as the InceptionV2 architecture is capable of learning more complex features. It is clear from the results that InceptionV2 is a good fit for the research domain. Due to its capacity to learn more complicated information, the InceptionV2 architecture is appropriate for categorising Handwriting from learners with and without down syndrome. The InceptionV2 architecture is the best architecture for the field of categorising Handwriting as down syndrome or not, according to the accuracy, loss, confidence score, precision, recall, and f1 score.

The Ecruite software was used in the study by *Tsao et al.* to identify the spatiotemporal indices that influence how learners with down syndrome write (Tsao et al., 2017). The researchers concluded that there is a developmental delay in writing acquisition for learners and adults with down syndrome. This research paper demonstrates how a deep learning architecture can identify whether an image is a down syndrome or non-down syndrome learner's Handwriting, which can subsequently help in providing feedback to learners with down syndrome on places to improve.

7 CONCLUSIONS

Although deep learning has been around for some time (Schmidhuber, 2015), deep learning has seen limited contributions within the down-syndrome Handwriting domain. This research paper introduced the difficulty of manually classifying and capturing down syndrome handwriting. The manual approach required much time and was error-prone. This study used an end-to-end deep learning model to differentiate handwritten images as down syndrome or non-down syndrome using three deep learning architectures, VGG16, InceptionV2, and Xception. This research paper introduced a new dataset for down syndrome handwriting while also examining the configurations of each architecture, outlining the advantages of transfer learning and looking at the results. Despite the VGG16 architecture's impressive performance in the research paper, more recent architectures like the IncpetionV2 and Xception perform better. The InceptionV2 could classify a single image with a confidence rating of 99.62% and an average f1-score of 97.10%. Similarly, the InceptionV2 architecture was able to classify each sample set with high accuracy, with only 3 false positives and an accuracy of 99.56%. InceptionV2 has outperformed the VGG16 and Xception architectures making it the ideal architecture for classifying the Handwriting of down-syndrome learners. The InceptionV2 architecture addresses the problem of the poor handwriting feedback provided to learners with down syndrome. The research paper advances knowledge of the characteristics of Handwriting that can be utilised to distinguish between down syndrome and non-down syndrome handwriting. Future research will focus on developing a more concrete tool to assess the level of readability of down syndrome learners handwriting, which will yield more informative results by looking at different classes. In conclusion, the InceptionV2 architecture can be used as a faster, more efficient way to distinguish between down syndrome and non-down syndrome handwriting. This solution can be a way to encourage more in-depth analysis to produce more accurate results for the recognition of down syndrome handwriting.

REFERENCES

Adamu, I., Umar, K. I., Watsila, H., and Ibrahim, B. M. (2017). Multi-layer perceptron network for english character recognition. *American Journal of Engineering Research (AJER)*, 6(6):86–92.

Agarwal, M. (2017). Understanding and coding inception module in keras.

Akhtar, Z. (2021). Xception: Deep learning with depthwise separable convolutions.

Almalaq, A. and Edwards, G. (2017). A review of deep learning methods applied on load forecasting. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 511–516. IEEE.

Anwar, A. (2019). Difference between alexnet, vggnet, resnet and inception. *Medium-Towards Data Science*.

Association, D. S. (2017). May's lp focus.

Brownlee, J. (2019). How to fix the vanishing gradients problem using the relu. *Machine Learning Mastery Website.*

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807.

Chumlea, W. C., Malina, R. M., Rarick, G. L., and Seefeldt, V. D. (1979). Growth of short bones of the hand in children with down's syndrome. *J. Ment. Defic. Res.*, 23(2):137–150.

Eaton, W. O., Bodnarchuk, J. L., and McKeen, N. A. (2014). Measuring developmental differences with an age-of-attainment method. *SAGE Open*, 4(2):2158244014529775.

Engel-Yeger, B., Nagauker-Yanuv, L., and Rosenblum, S. (2009). Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia. *The American Journal of Occupational Therapy*, 63(2):182–192.

Lee, A., Chung, J., and Lee, M. (2021). *GNHK: A Dataset for English Handwriting in the Wild*, pages 399–412.

Llamas, J., M. Lerones, P., Medina, R., Zalama, E., and Gómez-García-Bermejo, J. (2017). Classification of architectural heritage images using deep learning techniques. *Applied Sciences*, 7(10):992.

Muhammad, W., Aramvith, S., and Onoye, T. (2021). Multi-scale xception based depthwise separable convolution for single image super-resolution. *PLOS ONE*, 16(8).

Nguyen, L., Lin, D., Lin, Z., and Cao, J. (2018). Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation. pages 1–5.

Pandit, T., Kapoor, A., Shah, R., and Bhuva, R. (2020). Understanding inception network architecture for image classification.

Patton, S. and Hutton, E. (2017). Exploring the participation of children with down syndrome in handwriting without tears. *Journal of Occupational Therapy, Schools, & Early Intervention*, 10(2):171–184.

Priatama, A. R., , and Setiawan, Y. (2022). Regression models for estimating aboveground biomass and stand volume using landsat-based indices in post-mining area. *Jurnal Manajemen Hutan Tropika (Journal of Tropical Forest Management)*, 28(1):1–14.

Puranik, C. S., Petscher, Y., and Lonigan, C. J. (2013). Dimensionality and reliability of letter writing in 3-to 5-year-old preschool children. *Learning and Individual Differences*, 28:133–141.

Robert, Ross, M. E. G. A. and Thomas. Papers with code - imagenet benchmark (image classification). *The latest in Machine Learning*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

Shu (2019). Deep learning for image classification on very small datasets using transfer learning. 345.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision.

Theckedath, D. and Sedamkar, R. (2020). Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science*, 1(2):1–7.

Tsao, R., Moy, E., Velay, J.-L., Carvalho, N., and Tardif, C. (2017). Handwriting in children and adults with down syndrome: Developmental delay or specific features? *American Journal on Intellectual and Developmental Disabilities*, 122(4):342–353.

Wani, M. A., Bhat, F. A., Afzal, S., and Khan, A. I. (2020). Basics of supervised deep learning. In *Advances in Deep Learning*, pages 13–29. Springer.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168:022022.

Yu, W., Yang, K., Bai, Y., Xiao, T., Yao, H., and Rui, Y. (2016). Visualizing and comparing alexnet and vgg using deconvolutional layers. In *Proceedings of the 33 rd International Conference on Machine Learning*.