JOURNAL OF
EMERGING INVESTIGATORS

Article

Using two-stage deep learning to assist the visually impaired with currency differentiation

Krish Nachnani[1], Mikhil Kiran[2], Nisha Talagala[2]
[1] The Harker School, San Jose, California
[2] AIClub Research Institute, Saratoga, California

SUMMARY

Approximately 2.2 billion individuals across the world possess varying degrees of visual impairments, and they are challenged with a plethora of difficulties in their day-to-day experiences. In this study, we explored the application of deep learning algorithms to aid individuals with visual impairments in recognizing United States (US) currency. To address real-world issues such as damaged bills and background variations, we created a custom dataset specifically tailored for this task. We hypothesized that the utilization of a two-stage deep learning model would be more efficient and accurate than a single-stage deep learning model to address this problem. To test this hypothesis, we used the aforementioned artificial intelligence (AI) models trained on a custom dataset specifically designed to address real-world challenges to solve this problem. We chose the MobileNetV2 architecture for our experiments owing to its small size and potential for deployment on a mobile device. We also analyzed the importance of choosing optimal model hyperparameters in the context of this dataset. Through our experiments and choosing the models with the best validation accuracies, we achieved a test accuracy of 89% with a single-stage AI and a test accuracy of 83% with a two-stage AI. Our results, however, showed that a multi-level deep learning model did not provide any significant advantage over a single-level AI.

INTRODUCTION

Approximately 2.2 billion individuals across the world face challenges in their daily lives, as they possess varying degrees of visual impairments. Dealing with the loss of sight is challenging in itself, not to mention a lack of accessibility to resources, a societal stigma, and increased unemployment (1). Studies estimate that the burden on the economy by the visually impaired (direct medical costs and productivity losses) is approximately $16,838 annually per person (2, 3).

Visually impaired individuals in the United States face challenges when it comes to distinguishing currency. The absence of distinctive physical features like such as discernible crevices, size variations between bills makes it impossible for them to identify bills. The inability to independently identify currency puts them at risk of exploitation and having to rely on assistance from others to manage currency. We aimed to address the challenge of currency differentiation for individuals with visual impairments in our study (4).

To solve this problem, we employed deep learning models to recognize the differences between dollar bills. The first approach we implemented was a single-stage artificial intelligence (AI) model, which processes the image through a single AI model. After manually analyzing our dataset, we realized that a two-stage AI model, which processes the image through two different AI models before giving an output, might be a better solution.

We hypothesized that a two-stage image classification AI would work better than a single-stage image classification AI in addressing the aforementioned currency differentiation challenge. To test this hypothesis, we devised a comprehensive dataset of currency images to train the two different AI models and tested their accuracies.

RESULTS

We constructed a training data set that deliberately encompasses currency images captured from various angles, encompassing diverse rotations, different levels of damage, and even images with colorful backgrounds. These specific characteristics are important as they help with the optimization of the algorithm for usability in real-world scenarios.

We first employed a deep learning model to classify six types of bills($1, $5, $10, $20, $50 and $100), considering each bill as having two categories for the front and back. This results in a total of 12 categories. We call this approach the single-stage artificial intelligence (AI). However, since we observed that our dataset comprised various types of bill images, including close-ups, background variations, and rotations, we designed a different architecture: an AI model to determine the type of image, distinguishing between close-ups, images with background, and rotated images. After this, we trained 3 separate AI models for each bill category, where each model classified the bills into 1 of the 12 image categories. We call this architecture, which needs two stages of prediction and a total of four models, the two-stage AI.

In the development of our AI model, we employed the MobileNetV2 algorithm, a powerful architecture designed for complex image recognition tasks. This algorithm is readily accessible as a pre-trained network in the TensorFlow library. To achieve optimal performance, we adopted a technique known as transfer learning, which leverages the knowledge acquired from previous analyses to enhance the current classification task.

In our case, we utilized a version of the AI model that had undergone training on the widely used ImageNet dataset. This pre-training step enabled our AI model to use the general image recognition concepts effectively. By using transfer learning from the pre-trained MobileNetV2 model, we utilized the previous knowledge of the algorithm on a different large dataset to address the specific problem of currency differentiation. This approach significantly streamlined the

https://doi.org/10.59720/23-045

training process and resulted in the high accuracy of our AI model.

To improve our algorithm's peak accuracy, we performed numerous experiments by tweaking two hyper-parameters: epochs, and learning rate (5). Hyper-parameters are preset values that directly affect the speed and quality of the AI's training. Epochs are the number of times that an algorithm loops through the complete training dataset. Learning rate influences the rate at which new information replaces old information, essentially representing the speed at which a machine-learning model learns.

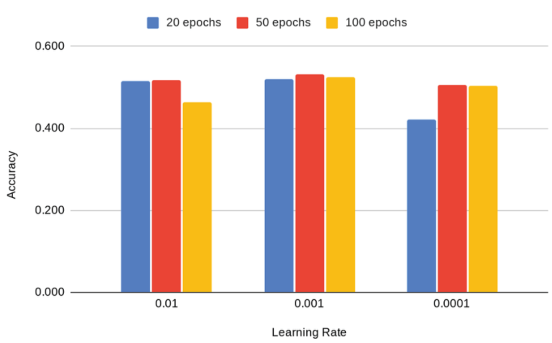
Single-Stage AI

For our single-stage AI model, we ran experiments with 20, 50, and 100 epochs. For each epoch, we changed the learning rate to 0.0001, 0.001, and 0.01. The validation accuracy ranged from 42% to 53%. The best validation accuracy was 53% and was obtained for a learning rate of 0.001 and an epoch value of 50 **(Figure 1A)**. The confusion matrix indicates that the model was confused the most between the back of a \$20 bill and the back of a \$50 bill **(Figure 1B)**. The classification report showing the precision, recall, and F1 scores for each category shows that the front of a \$50 bill had the lowest precision of 0.0 and the lowest recall value of 0.0. The front of a \$10 bill, on the other hand, had the best precision of 1.0, and the front of a \$20 bill had the best recall value of 1.0 **(Figure 1C)**. The test accuracy of the model on a separate test dataset was 89%. The confusion matrix of the test dataset indicates that the model was confused the most between the back of a \$20 bill and the back of a \$50 bill **(Figure 2A)**.

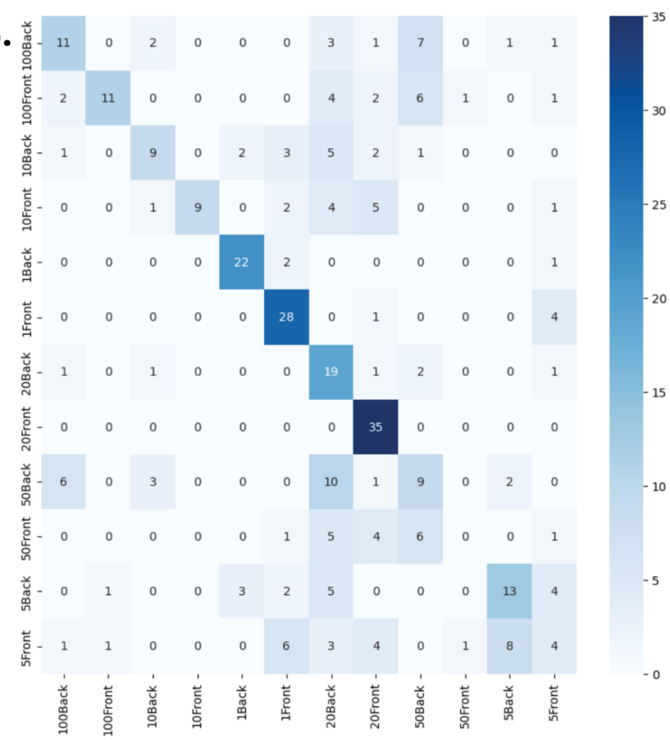
We have 4 different AI models that together make up the two stage AI architecture - Selector, Closeup, With Background, and Rotated. We ran experiments with 20, 50, and 100 epochs and changed the learning rates to 0.0001, 0.001, and 0.01 for each epoch. The four models exhibited different average validation accuracies for all epochs and learning rates: the selector AI model had validation accuracies between 88%-90%, the accuracies for the closeup AI were between 78%-95%, the with background between 29%-41%, and for the rotated it was between 76%-89% **(Figures 3A, 4A, 5A, 6A)**. The AI models also showed high confusion rates with different categories, as the selector was confused the most between categories 'closeup' and 'rotated' **(Figure 3B)**. The other three AI models were all confused on different categories, with the closeup being the most confused between the front of a \$5 dollar bill and the front of a \$1 dollar bill, the with background model being confused the most between the front side of both a \$5 dollar bill and a \$10 dollar bill, and the rotated model being the most confused between the front of a \$20 bill and the front of a \$10 bill, and the front of a \$20 bill and the front of a \$5 bill **(Figures 4B, 5B, 6B)**.

Using data from the classification reports that show the precision, recall, and F1 scores for each category, we were able to discern which categories the AI models had the most success and most failure with. The data for the selector model shows that 'closeup' had the lowest precision of 0.89 and 'rotated' had the lowest recall value of 0.87, while 'rotated' had the best precision value of 0.9 and 'with background' had the best recall value of 1.00 **(Figure 3C)**. The classification

A.



B.



C.

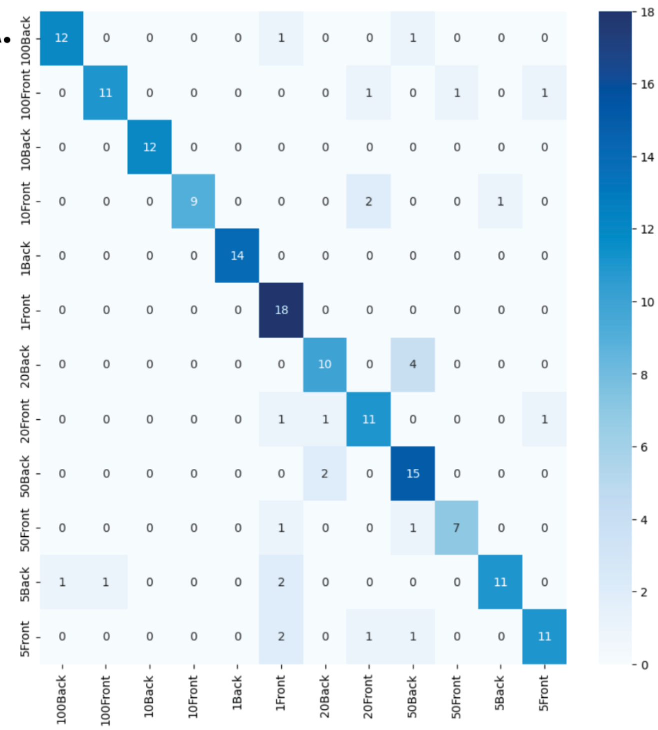
Image Category	Precision	Recall	F1-Score	Support
100back	0.50	0.42	0.46	26
100front	0.85	0.41	0.55	27
10back	0.56	0.39	0.46	23
10front	1.00	0.41	0.58	22
1back	0.81	0.88	0.85	25
1front	0.64	0.85	0.73	33
20back	0.33	0.76	0.46	25
20front	0.62	1.00	0.77	35
50back	0.29	0.29	0.29	31
50front	0.00	0.00	0.00	17
5back	0.54	0.46	0.50	28
5front	0.22	0.14	0.17	28

Figure 1: Validation results for the Single-Stage AI. A) An Accuracy vs. Learning Rate graph for 20, 50, and 100 epochs. B) A confusion matrix for the validation accuracies. C) A classification report for the validation accuracies.

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/23-045

A.



B.

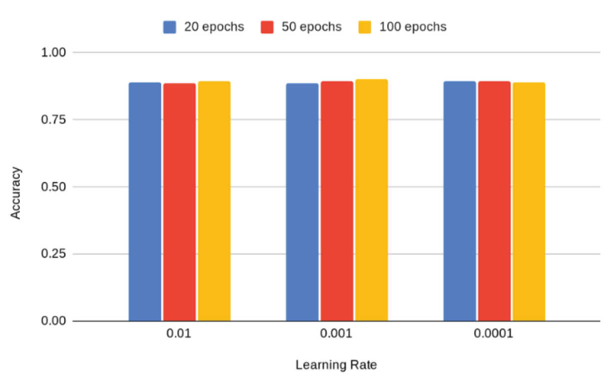
Image Category	Precision	Recall	F1-Score	Support
100back	1.00	1.00	1.00	9
100front	0.69	1.00	0.82	9
10back	0.88	0.78	0.82	9
10front	1.00	1.00	1.00	9
1back	0.90	1.00	0.95	9
1front	1.00	1.00	1.00	3
20back	1.00	0.89	0.94	9
20front	1.00	0.11	0.20	9
50back	1.00	1.00	1.00	9
50front	0.90	1.00	0.95	9
5back	1.00	1.00	1.00	9
5front	0.69	1.00	0.82	9

Figure 2: Test results for the Single-Stage AI. A) A confusion matrix for the test accuracies. B) A classification report for the test accuracies.

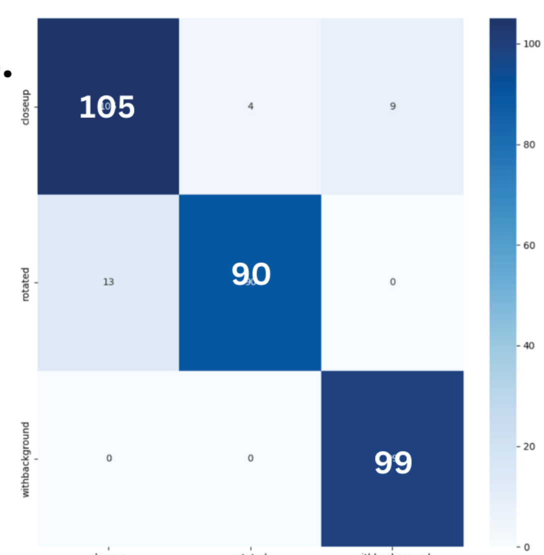
report for the closeup model shows that the front of a $1 bill had the lowest precision of 0.85 and the front of a $5 bill had the lowest recall value of 0.78. The back of a $100 bill, the front of a $100 bill, the back of a $10 bill, the front of a $10 bill, the back of a $1 bill, the back of a $50 bill, the front of a $50 bill, the back of a $5 bill, and the front of a $5 bill, all had the best precision value of 1.00. The front of a $100 bill, the back of a $10 bill, the front of a $10 bill, the back of a $1 bill, the front of a $1 bill, the front of a $20 bill, the back of a $50 bill, and the back of a $5 bill all had the best recall value of 1.00 **(Figure 4C)**.

The classification report for the with background model shows that the front of a $100 bill and the back of a $5 bill both had the lowest precision of 0.00 and the lowest recall value of 0.00. The back of a $1 bill, on the other hand, had the best precision of 0.75 and also the highest recall value

A.



B.



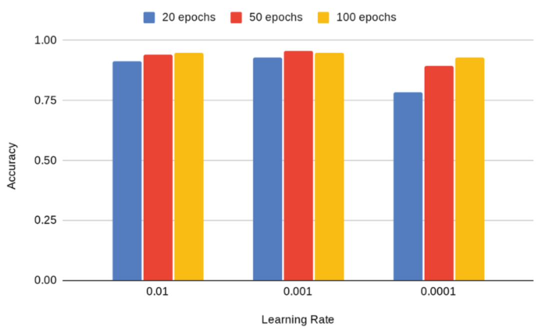
C.

Image Category	Precision	Recall	F1-Score	Support
closeup	0.89	0.89	0.89	118
rotated	0.96	0.87	0.91	103
withbackground	0.92	1.00	0.96	99

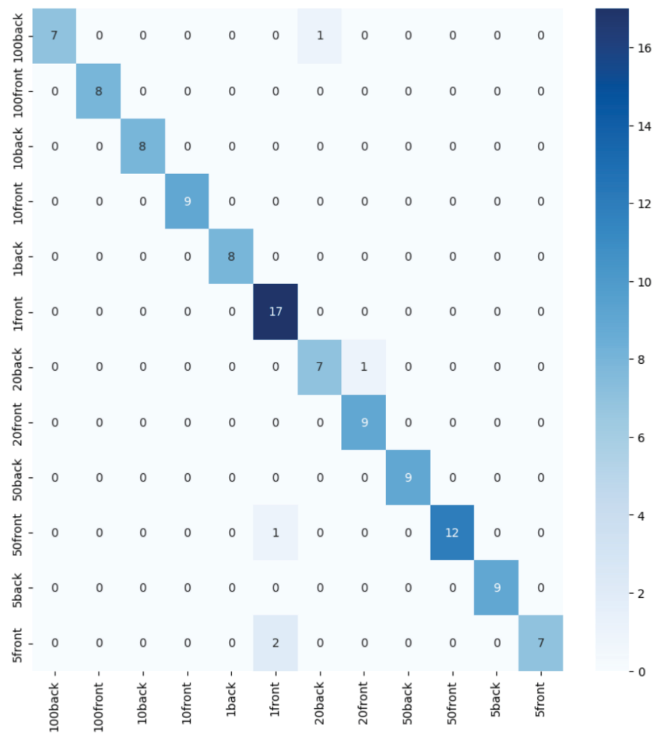
Figure 3: Validation results for the Selector AI. A) An Accuracy vs. Learning Rate graph for 20, 50, and 100 epochs. B) A confusion matrix for the validation accuracies. C) A classification report for the validation accuracies.

https://doi.org/10.59720/23-045

A.



B.

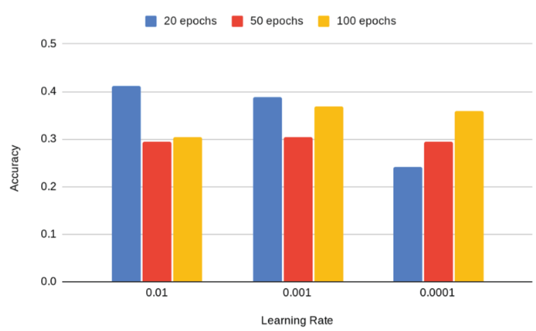


C.

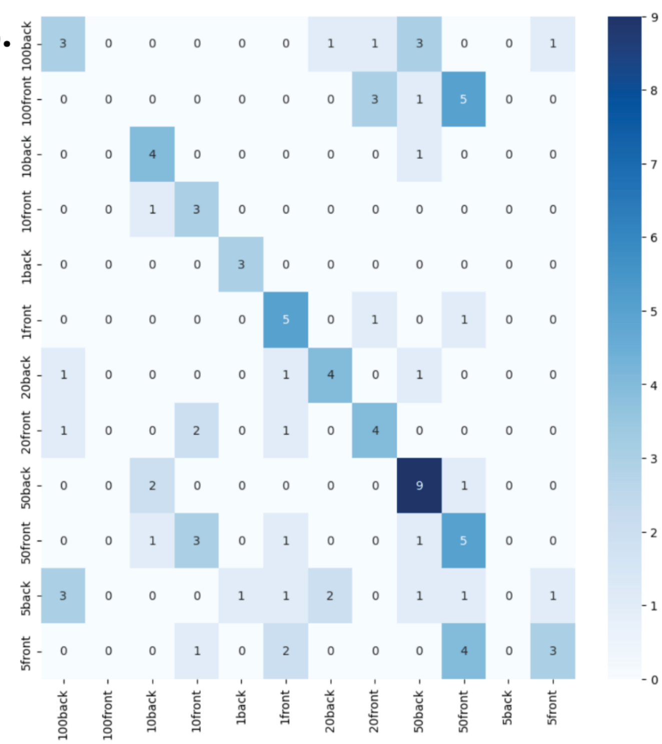
Image Category	Precision	Recall	F1-Score	Support
100back	1.00	0.88	0.93	8
100front	1.00	1.00	1.00	8
10back	1.00	1.00	1.00	8
10front	1.00	1.00	1.00	9
1back	1.00	1.00	1.00	8
1front	0.85	1.00	0.92	17
20back	0.88	0.88	0.88	8
20front	0.90	1.00	0.95	9
50back	1.00	1.00	1.00	9
50front	1.00	0.92	0.96	13
5back	1.00	1.00	1.00	9
5front	1.00	0.78	0.88	9

Figure 4: Validation results for the Closeup AI. A) An Accuracy vs. Learning Rate graph for 20, 50, and 100 epochs. B) A confusion matrix for the validation accuracies. C) A classification report for the validation accuracies.

A.



B.



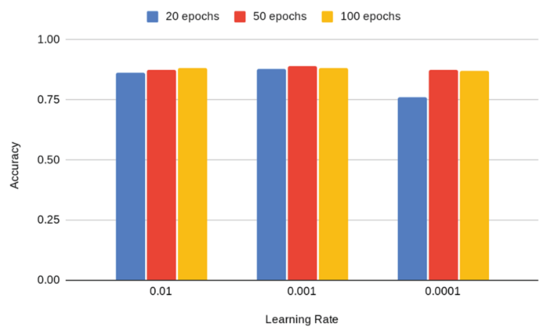
C.

Image Category	Precision	Recall	F1-Score	Support
100back	0.38	0.33	0.35	9
100front	0.00	0.00	0.00	9
10back	0.50	0.00	0.62	5
10front	0.33	0.75	0.46	4
1back	0.75	1.00	0.86	3
1front	0.45	0.71	0.56	7
20back	0.57	0.57	0.57	7
20front	0.44	0.50	0.47	8
50back	0.53	0.75	0.62	12
50front	0.29	0.45	0.36	11
5back	0.00	0.00	0.00	10
5front	0.69	0.30	0.40	10

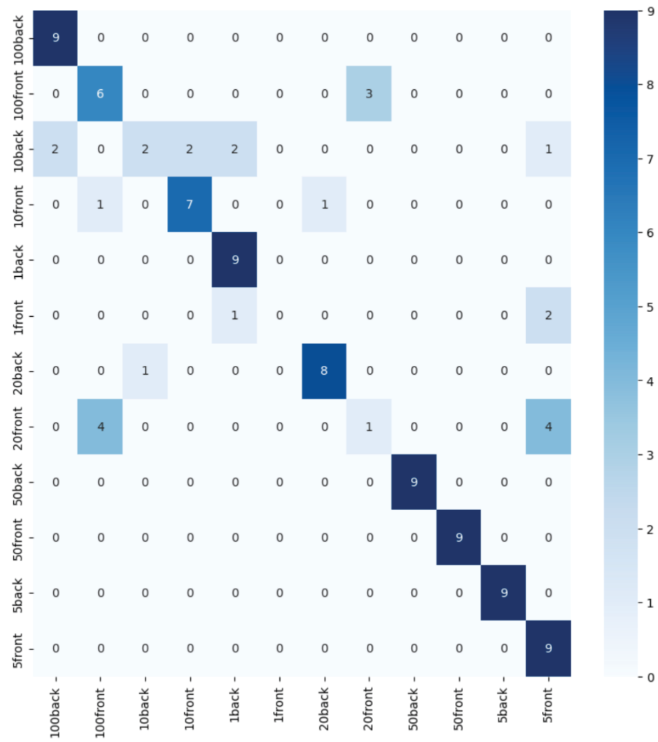
Figure 5: Validation results for the With Background AI. A) An Accuracy vs. Learning Rate graph for 20, 50, and 100 epochs. B) A confusion matrix for the validation accuracies. C) A classification report for the validation accuracies.

https://doi.org/10.59720/23-045

A.



B.



C.

Image Category	Precision	Recall	F1-Score	Support
100back	1.00	1.00	1.00	9
100front	0.82	1.00	0.90	9
10back	1.00	0.78	0.88	9
10front	1.00	1.00	1.00	9
1back	0.82	1.00	0.90	9
1front	0.67	0.67	0.67	3
20back	1.00	1.00	1.00	9
20front	1.00	0.11	0.20	9
50back	1.00	1.00	1.00	9
50front	1.00	1.00	1.00	9
5back	0.90	1.00	0.95	9
5front	0.64	1.00	0.78	9

Figure 6: Validation results for the Rotated AI. A) An Accuracy vs. Learning Rate graph for 20, 50, and 100 epochs. B) A confusion matrix for the validation accuracies. C) A classification report for the validation accuracies.

of 1.00 **(Figure 5C)**. Finally, the classification report for the rotated model shows that the front of a $1 bill had the lowest precision of 0.67 and the front of a $20 bill had the lowest recall value of 0.11. The back of a $100 bill, the back of a $10 bill, the front of a $10 bill, the back of a $20 bill, the front of a $20 bill, the back of a $50 bill, and the front of a $50 bill all had the best precision values of 1.00. It is also worth noting that the recall value is equal to 1.00 for 9 out of the 12 categories. **(Figure 6C)**.

Two-Stage AI

For the test accuracy of the two-stage AI, we ran the test dataset through the most accurate hyperparameter combination for the selector AI. Based on the output category, we ran the test dataset through the rotated, with background, or closeup AI models using the most accurate hyperparameter combination. The test accuracy of the two-stage AI model was 83%. The confusion matrix of the test dataset indicates that the model was confused the most between the back of a $5 bill and the front of a $5 bill **(Figure 7A)**.

DISCUSSION

The two main contributions of our work include the creation of a hand-assembled dataset tailored to address real-world challenges faced by individuals with visual impairments and the development and evaluation of deep learning techniques. Our dataset accounts for various scenarios that ultimately allow for usability in real-life situations, including differentiating between the front and back of currency and crumpled, folded, or creased bills. We explored two approaches in using deep learning models, where we conducted experiments to determine the impact of various hyperparameters on the predictive performance of the model. For the single-stage AI approach, we achieved a test accuracy of 89% where the training model was created with a learning rate of 0.001 and 50 epochs. The test accuracy of the two-stage AI, on the other hand, was 83%.

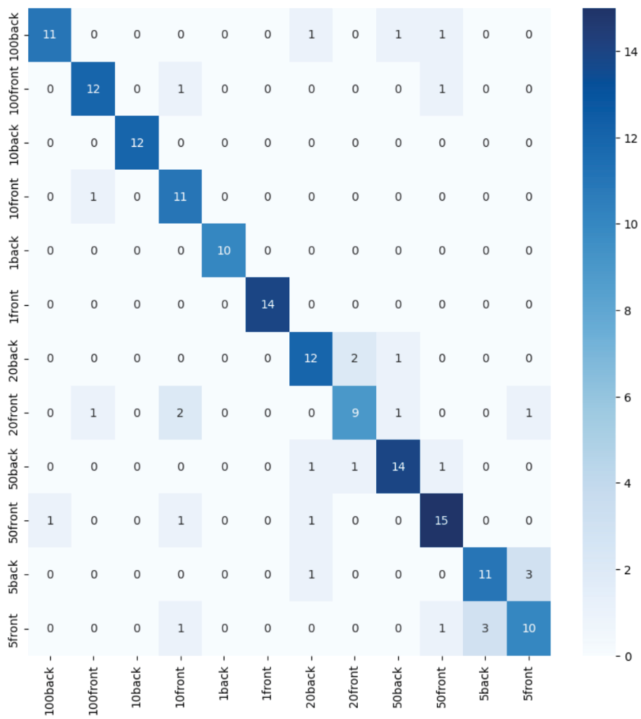
We did not notice a general trend of either an increase or decrease in validation accuracy based on the learning rate or the epoch values. However, in some specific cases, we noticed that the validation accuracy improved with an increase in epoch value. This happened in the case of the Single Level AI when the learning rate was 0.0001 **(Figure 1A)**, the Close Up AI when the learning rate was 0.01 and 0.0001 **(Figure 4A)**, and the With Background AI when the learning rate was 0.0001 **(Figure 5A)**.

The hyper-parameter tuning graphs show that increasing the epoch value from 50 to 100 did not help in improving the accuracy value in general **(Figures 1A, 3A, 4A, 6A)**. We also noticed that in the case of the With Background model, when the epoch value was 20, the validation accuracy increased with an increase in learning rates, while in the case of 100 epochs for the same AI model, the validation accuracy decreased with an increase in learning rate **(Figure 5A)**. In general, the results indicate that tweaking the epochs has much less impact on the validation accuracies compared to the learning rate.

In our study, we were able to effectively compare and contrast our two AI models, a single-stage and two-stage AI. We discovered that a single-stage AI system is generally more effective at detecting and classifying currency under various real-life situations. Typically, an ensemble of weak

https://doi.org/10.59720/23-045

A.



B.

Image Category	Precision	Recall	F1-Score	Support
100back	0.92	0.79	0.85	14
100front	0.86	0.86	0.86	14
10back	1.00	1.00	1.00	12
10front	0.69	0.92	0.79	12
1back	1.00	1.00	1.00	10
1front	1.00	1.00	1.00	14
20back	0.75	0.80	0.77	15
20front	0.75	0.64	0.69	14
50back	0.82	0.82	0.82	17
50front	0.79	0.83	0.81	18
5back	0.79	0.73	0.76	15
5front	0.71	0.67	0.69	15

Figure 7: Test results for the Two-Stage AI. A) A confusion matrix for the test accuracies. B) A classification report for the test accuracies.

classifiers results in a stronger classifier. In the ensemble method, the classifiers predict the same outcome, and their predictions are pooled to make the final prediction. However, in our case, the classifiers are used in tandem rather than as an ensemble. When the classifiers are used in tandem, the mispredictions of the first classifier seem to be magnified by the second classifier, as they solve two different tasks. This could be the reason that the two-stage AI did not have better performance than the single-stage model.

The test results indicated that for a single-stage AI, the front of a $100 bill and the front of a $5 bill had the worst precision value **(Figure 2B)**. In the case of a two-stage AI, the front of a $10 bill had the worst precision value. **(Figure 7B)**. It is interesting to see how the two approaches differ in the categories where the model has difficulty identifying a category. This could be the result of the two-stage AI pipeline having to deal with a smaller number of categories compared to the single-stage AI. These findings emphasize the influence of pipeline design on the models' abilities to distinguish between different categories, despite utilizing the same MobileNetV2 architecture.

Our results show that our hypothesis is not validated by the results from the experiments. A two-stage AI system performs worse than a single-stage AI. In the past, there has been minimal evidence of an AI with multiple layers developed to solve a problem. While the basis of the AI may contain many layers of deep learning algorithms, the problem itself is not solved with multiple AIs. We believe that this project is one of the few that tested this strategy. Overall, our study contributes to the understanding of currency recognition using deep learning models and provides insights for the development of effective AI systems in real-world applications. Such advances will enable visually impaired individuals to use physical currency without fear of being shortchanged and enable them to be less reliant on others.

MATERIALS AND METHODS

The Dataset

The dataset comprised a total of 1,788 images, predominantly obtained through manually taking pictures of various currency notes, with a few exceptions sourced from the internet **(Figure 8)**. Specifically, the dataset encompassed various denominations of dollar bills, including $1, $5, $10, $20, $50, and $100.

In order to train the model using the two-stage AI, the images within the dataset were categorized into three distinct groups: (a) close-up images, (b) images with colorful backgrounds, and (c) images rotated in increments of 15 degrees. We split the dataset up into these groups to mimic variations in real-world scenarios. Within each group, there were 12 categories of images, as we used 6 different dollar bills, and we took pictures of the front and back sides of each bill. These 12 categories were the front of $100 bill, back of $100 bill, front of $50 bill, back of $50 bill, front of $20 bill, back of $20 bill, front of $10 bill, back of $10 bill, front of $5 bill, back of $5 bill, front of $1 bill, and back of $1 bill. We used this organized dataset when training the two-stage AI.

We split the dataset into separate train and test sets. We then verified that both the train and test sets had similar distributions of images across all categories. We further divided the training dataset into train and validation subsets during hyper-parameter tuning in our experiments.

Experimental Design

We created two variations of AI: a single-stage AI and a two-stage AI - a type of ensemble approach (6). In the single-

https://doi.org/10.59720/23-045

Figure 8: Dataset Image Examples. The images include photos that are taken from close up, rotated photos, and photos with vibrant backgrounds.

stage AI, we combined the three categories of datasets, which we acquired through taking numerous pictures, into a singular dataset with some images from each category, optimizing one AI for all categories. For the two-stage AI, we created a selector AI to figure out the category of the image, then the second stage used the respective category dataset to make its prediction.

MobileNetV2

In creating the models, we utilized an AI algorithm named MobileNetV2. MobileNetV2 is a recent, state-of-the-art image recognition neural network library (a type of Convolutional Neural Network), which is the reason we determined that it was the perfect choice for our project (7, 8). Google's algorithm is optimized for small devices, such as portable computers, which allows everyday individuals to make use of it without splurging on the latest technology. Along with these positives, MobileNetV2's algorithm is optimized to conserve computing resources on devices, resulting in a minuscule environmental impact.

Transfer Learning

Before inputting our data into MobileNetV2, we trained a MobileNetV2 model on ImageNet, which is a library consisting of 14 million images separated into 20,000 categories. The categories include common words such as "balloon" or "strawberry," which each contain thousands of images that AIs can use to train (11). This dataset was the foundation for our AI's training on our dataset. All the MobileNetV2 models that were generated used transfer learning as a starting point (12). The model weights of the pre-trained model were frozen. A trainable dense layer was added on top of the pre-trained network to ensure that the model retains the knowledge from its training on the ImageNet dataset while learning new information needed using the dense layer that was added on top of the pre-trained network.

Received: October 24, 2022
Accepted: August 19, 2023
Published: June 2, 2024

REFERENCES

1. "Vision Impairment and Blindness." *World Health Organization,* www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. Accessed 1 Oct. 2022.
2. "The Economic Burden of Vision Loss and Eye Disorders in the United States." *University of Chicago,* www.norc.org/Research/Projects/Pages/the-economic-burden-of-vision-loss-and-eye-disorders-in-the-united-states.aspx. Accessed 1 Oct. 2022.
3. "The Economic Burden of Vision Loss and Eye Disorders in the United States." CDC, www.cdc.gov/visionhealth/projects/economic_studies.htm. Accessed 1 Oct. 2022.
4. National Research Council. *Currency features for visually impaired people.* Vol. 478. National Academies Press, 1995.
5. Battini, Deepak. "Basic Hyperparameter Tuning for Neural Networks." *Tech,* www.tech-quantum.com/basic-hyperparameter-tuning-for-neural-networks. Accessed 1 Oct., 2022.
6. "Ensemble Learning." *Wikipedia,* en.wikipedia.org/wiki/Ensemble_learning. Accessed 1 Oct., 2022.
7. Saha, Sumit. "A Comprehensive Guide to Convolutional

https://doi.org/10.59720/23-045

Neural Networks-the eli5 Way.” *Medium,* Towards Data Science, 17 Dec. 2018, towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. Accessed 1 Oct, 2022.
8. Sandler, Mark, *et al.* “MobileNetV2: Inverted Residuals and Linear Bottlenecks.” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 21 Mar. 2019, https://doi.org/10.1109/cvpr.2018.00474.
9. He, Kaiming, *et al.* “Deep Residual Learning for Image Recognition.” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 10 Dec. 2015, https://doi.org/10.1109/CVPR.2016.90.
10. Tan, Mingxing, and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” 2019 *International Conference on Machine Learning,* 11 Sept. 2020, https://doi.org/10.48550/arXiv.1905.11946.
11. “From Not Working to Neural Networking.” *The Economist,* www.economist.com/special-report/2016/06/23/from-not-working-to-neural-networking. Accessed 1 Oct. 2022.
12. “Transfer Learning and Fine-Tuning: Tensorflow Core.” *TensorFlow,* www.tensorflow.org/tutorials/images/transfer_learning. Accessed 1 Oct. 2022.