Detecting Cognitive Decline Using a Novel Doodle-Based Neural Network

1st Connor Pearson
Computing Science Department
University of East Anglia
Norwich, England
zvu19cpu@uea.ac.uk

2nd Beatrice de la Iglesia
Computing Science Department
University of East Anglia
Norwich, England
b.iglesia@uea.ac.uk

3rd Saber Sami
Norwich Medical School
University of East Anglia
Norwich, England
s.sami@uea.ac.uk

***Abstract*—A key part in the diagnosis of cognitive decline are visuospatial based tests. These visuospatial tests often involve a form of drawing task. In this paper, we build an automated multi-class classifier to assign hand-drawn doodles from Google's online game Quick, Draw! into 24 unique categories that are simple to draw doodles. Our goal is to create a prototype of an automated online diagnosis tool that resembles the visuospatial portion of established pen and paper cognitive examinations. We built a CNN using the Tensor Flow Keras API, and tested multiple iterations of each model neuron structure. We created a web interface able to capture user inputs from a browser window as they draw the requested doodle for each test stage. The images are relayed back to a server and processed through the same model trained on the Google QuickDraw! dataset to determine a patient's score. Herein we use these model predictions as a measurement of the users drawing skills. Using a CNN based neural network we achieved a 90.46% model accuracy and around 70% implementation accuracy which is not dissimilar to human pen and paper ratings.**

I. Introduction

Through the automation of medical testing we benefit from easier access to regular tests which "saves time and enable(s) timely interventions over complex cases" [1]. We're even seeing increases in testing accuracy from online tests.

Additionally, with the reduced cost to testing, machine learning solutions can be used in parity with conventional testing to prevent the misdiagnosis of patients without much additional cost. With certain tests, such as some neurological tests, we can make solutions that can be done from home via a smart device [2].

Dementia is a difficult disease to diagnose, especially with the limited pen and paper tools available to us in the past.

The aim here is to create a automated online implementation of a well documented dementia diagnosis test. Using the Addenbrooke's Cognitive Examination [3] as a foundation the aim is to digitise the visuospatial questions and automate the processing of results.

II. Aims

The project aims to provide an accurate online testing solution resembling the pen and paper ACE III test which utilises known "neurobiological underpinnings" present in those who suffer from early stage dementia [4] .

The solution aims to be accompanied by an intuitive User Interface which considers the users potential disabilities and strongly follows accessibility standards.

A. Objectives

1) Produce an accurate Machine Learning model capable of processing a users answers to provide a diagnosis in the form of a numerical score.
   This is measurable from the accuracy we achieve in our solution.
2) Produce an intuitive web interface which abides by the standard HTML Access-ability standards.
   This would be measurable by the consistency of the webpage design and processing of HTML source code in online HTML validation checkers.
3) Produce an web based solution with account functionality capable of storing a user's results and potentially save a users progress.
   This would be measurable in the solutions ability to provide account creation and logging in functionality.

III. Design

A. Initial design planning

1) Issues and Risks:

a) Classifying natural decline against neurological diseases: A potential use of the program would be the classification of cognitive decline rate into two categories. Determining if the current rate of decline (derived from testing) is simply correlated to the patients increased ageing, or if the decline is representative of a potential case of Dementia.

This computational approach help derive the diagnosis as there may be deviance in the rate of natural decline in individuals on a case-by-case basis. Meaning there could be very little difference between a patient with significant natural decline and a patient with mild cognitive deterioration indicative of dementia.

The ease of capturing online data could provide a solution comparing large cohort data of 'normal healthy' cognitive decline rates in order to determine if a patient is deviating.

b) Classification Accuracy: Though the visuospatial skills are proven to be an indicator of cognitive health, it is only one of many potential indicators. So it may prove difficult to form a diagnosis from the chosen data set solely on its own.

Another risk to this work is our ability to accurately class patient cognitive health using the visuospatial data set we are deriving.

IV. Implementation

A. Data

We used the "Quick,Draw!"data set from Google, which contains 50 million drawings across 345 categories contributed by players of the game. The game calls on players to draw a specific image in less than 20 seconds. While the user draws, an advanced neural network tries to establish the object's category and its predictions evolve as the user adds more and more detail. The drawings were captured as timestamped vectors, tagged with metadata including what the player was asked to draw and in which country the player was located.The data can be browsed at https://quickdraw. withgoogle.com/data. The data-set is available in different formats. One of the format is as simplified drawings in 28x28 grayscale bitmap in numpy .npy format. The files can be loaded with np.load(). We use the simplified dataset with images in .npy format as it has already been pre-processed, and it is well adapted for webapplications.

Since the data is stored in the Google database in category-wise format, we combine all 24 categories into single data-frame and then apply different algorithms to train our models. There are more than 100,000 images per category available.

B. Machine Learning Model

In order to recognise doodles in an effective manner this work considered multiple potential forms of machine learning from Convolutional Neural Networks (also know as CNNs) to novice Neural Networks in the form of Multi-Layer Perceptron's (MLP).

Data augmentation, which is the generation of any flipped or slightly rotated images, is often required when training data is limited. However, we had over 100,000 images per category so, data augmentation was not implemented.

It is also worth noting there is a discrepancy in results accuracy, as accuracy is based on the testing stage of development. This means the data used is sourced from the remaining doodles in the dataset that were acquired from QuickDraw! for validation testing. Not the doodles generated by users in our solution. Implementation accuracy is addressed separately as 'implementation accuracy' for each tested model.

C. Web-page solution

In line with our statements on accessibility of testing, specifically in terms of remote access in the 'post-pandemic' world, an online testing platform was created. The solution intends to be intuitive and feature elements of similar medical testing tools with "interfaces and subtests" [5] to provide a good user experience.

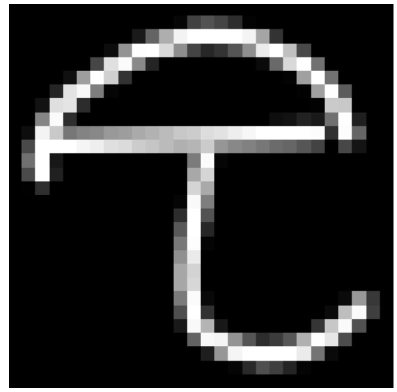

Fig. 1. Sample drawing of an umbrella from the Google QuickDraw! dataset used in the training of the Multi-layer Perceptron Model

Using a combination of JavaScript, CSS, HTML5 and PUG (a handy web-page template engine) a front-end was developed to provide users with an in-browser solution to take the test. Users navigate to the site which is running on a NodeJS based server with Python scripts used to process test images through our trained model and return results in the form of a JSON file.

The results from the server-side JSON are used to derive a numerical score for the client. This score will be indicative of the current cognitive health of the user.

D. Server-side implementation

1) NodeJS Server: The NodeJS server is based on a form of JavaScript and allows us to outline the ways in which web-pages and files are served to and received from the client. The server also manages sessions and appropriate file I/O operations.

In this implementation the NodeJS server plays a crucial role in the rendering of web-pages as it takes the PUG template web-page files and processes the data in the form of HTML to the front end.

Additionally, the server manages session directories, a session is generated client-side and is comprised of a Unix Epoch time string in milliseconds appended to a randomly generated UUID. This session ID is stored in a standard form of browser cache.

The server also receives the full resolution doodle from the client-side in the form of a 400 by 400 pixel image transferred from the client side inside a JSON package in the form of a Data-URL. The image data is then extracted and saved as a PNG in the generated session folder.

Once the test is complete the server opens a python shell and runs the image_processor.py script which takes the sessions PNG doodles and down-scales them to the 28 by 28 pixel doodles required to fit the Neural Networks model inputs. Following this step the server runs the query_model.py script. This scripts passes the users doodles through the model and saves the results in a JSON which is then passed back to the client to show the test score.

As the python scripts take a short time to compute there is a waiting page which pings the server to check if the python script is complete every 5 seconds. This page simply sends the current session ID to the server and the server checks for the existence of the results file containing the test results which only exists in the directory if the python scripts have finished rendering the results.

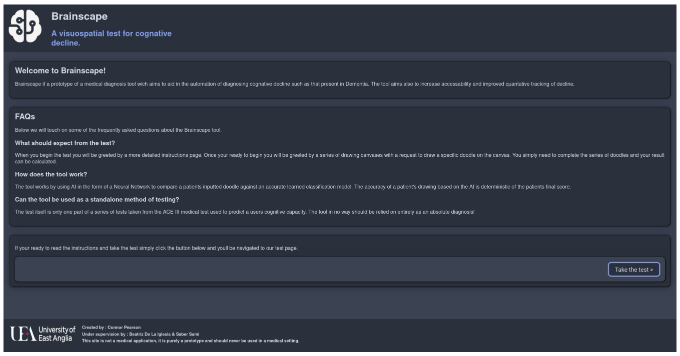

Fig. 2. Web-based visuospatial task homepage

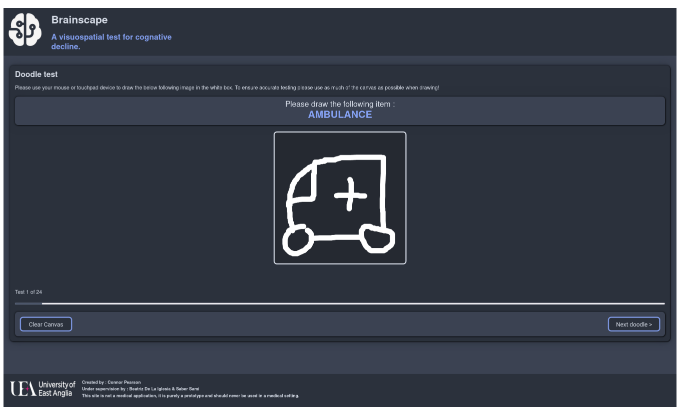

Fig. 3. User interface enabling dynamic free drawing

E. Analysis

In order to test the Neural Network models a set of batch jobs ran for several days on GPU accelerated servers. Each job took the Google QuickDraw! dataset and shuffled the data to ensure it was randomised. Once the data was randomised it was split into two sets one for training (representing 75% of the dataset) and one for testing (utilising the remaining 25% of the dataset). This means every job had a random assortment of training and testing data.

The jobs then processed a neural network for every combination of neuron structure and epoch count. So each job would run a model for all 5 epoch counts in the set of epochs 4, 6, 16, 32, 64 against all 4 of the set neuron structures in the set 128_16, 128_16_8, 256_32, 256_32_16, totalling a number of 20 separate models for each job.

In Table I we show the average accuracy values derived from the batch jobs. This was intended to allow us to account for any inaccuracies in our results due to the randomisation of datasets as well as to account for the inherent variance of Neural Networks.

F. Results

1) Model performance comparison: In the final web application we saw the Neural Network was capable of detecting user generated doodles.

The application provides a dynamic method of classifying doodles and provides features one of the main components of the ACE-III visuospatial test.

One unexpected result from the development of the program was the variance in accuracy between the models being tested against the training data and the models being tested against real world (user data). It seemed as though classification accuracy using the model with the highest testing accuracy (model ”256_32_16”) can vary significantly.

TABLE I
GRAPH SHOWING PERFORMANCE OF VARIOUS COMPETING MODELS WITH DIFFERING STRUCTURES AND EPOCH VALUES. IN BOLD WE CAN SEE OUR TWO BEST COMPETING RESULTS.

Model	**4 epoch**	**8 epoch**	**16 epoch**	**32 epoch**	**64 epoch**
128_16	0.7830	0.7304	0.7534	0.8777	0.8252
128_16_8	0.8148	0.8146	0.7811	0.8334	0.8637
256_32	0.8003	0.8168	**0.9088**	0.8718	0.8776
256_32_16	0.8280	0.8772	**0.9046**	0.8636	0.8226

Based on the Table I data we can dictate that when using the original QuickDraw! data for testing our highest average accuracies are with the model 256_32 at Epoch 16 and 256_32_16 at Epoch 16, this suggests that the model benefited greatly by the presence of denser neuron layers (additional neurons) in the structure.

2) Comparison of Results: As seen in the Fig. 4, models are plotted with the label representing the count of neurons in each layer. So the model shown to have the highest average accuracy across each training interval is labelled ”256_32_16” meaning the networks hidden layers consist of a layer of 256 neurons, then a layer of 32 neurons and finally a hidden layer with 16 neurons.

With the ”256_32_16” model we were able to achieve a theoretical testing accuracy as high as 90.46%.

Using user generated doodles we were achieving a user score of around 70 on some doodle sets of good drawing quality. This means the model was significantly better at classifying the testing set from the QuickDraw! dataset then the user generated doodles.

It was concluded the issue was likely some form of over training on the model ”256_32_16”. In order to resolve this a series of other models were tested in the program with lower epochs and a new model (model 4-256_32_16.h5) was chosen which returned an accuracy of around 70% indicated by the user scores it returned.

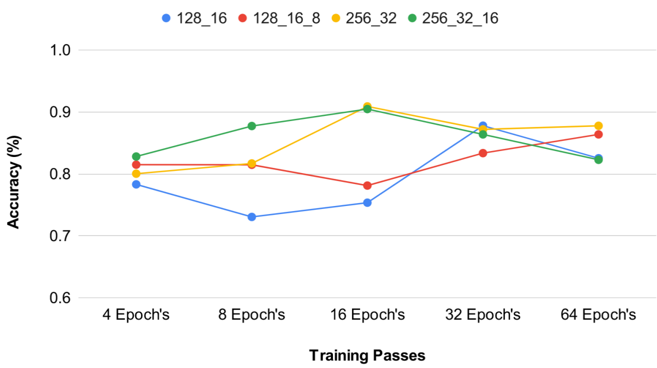

Fig. 4. Line graph showing the final average accuracy of the 4 chosen model structures at each epoch count. The models 256_32 at 16 epochs and 256_32_16 at 16 epochs represent the two highest accuracy models.

V. Evaluation

1) Discussion: Using a neural Network we managed to achieve a theoretical accuracy of 90.46% for our doodle recognition application in testing.

From the previous research we can determine that there are a large quantity of solutions aiming to provide digitised testing for people with Mild Cognitive Impairment and Alzheimer's Disease. However, a lot of these solutions do not meet the requirements to make them a valid and reliable form of testing due to a lack of sufficient results data.

We can also determine that there are many different methods of approaching a visuospatial test in the context of our study. For example, the quality and accuracy of a patients drawing is one method of measurement. But ideally we would use data from how the user draws a sketch, by logging the amount of time the user spends on drawing with a pen on screen, how long the patient spends thinking with the pen off the screen and also how long the user spends on the screen when creating pen strokes [6].

With this approach the intention is to automate the process for early diagnosis of Alzheimer's disease to a much greater accuracy. This approach not only saves time but increases accessibility to neurological tests while reducing demand for highly specialised individuals.

A. Future prospects

In this section we cover what future developments may be implemented to improve on the solution.

1) Account functionality in web application: A feature of interest that fell outside the scope of the study was account functionality to allow the storage of user scores over prolonged periods as well as an account hierarchy for medical professionals to access patient data and the trends their scores follow.

The dashboard would plot graphs and show what the status and trajectory of patients cognitive health looked like. This solution would likely require some form of SQL database implementation perhaps via MySQL to manage accounts and patients records more accurately and securely.

2) Smartphone integration: An important factor for future development could take inputs from a drawing pad or some form of smart device like a dedicated Android tablet or smartphone. This would act as a web container for the site and allow the user to draw on the testing canvas via the touchscreen.

B. Conclusions

In this work we found that the application of CNN/Multi Layer Perceptrons in the digitisation of a visuospatial test proves to be beneficial much like previous automation proposals [7]. The data returned from the work suggests that further development into a suite of testing tools specific to cognitive decline testing may very well prove effective. In addition expansion into newer more specialised technologies such as Convolutional Neural Networks (with the inclusion of transfer learning) may increase overall testing accuracy on specific tests.

In terms of accuracy of results, it should be noted that implementation accuracy needs further testing due to the fact we had a limited sample size, especially with Convolutional Neural Networks.

With further development put into the machine learning models a stronger solution capable of deriving the "information concerning psychometric quality" [8] of Dementia could be within our grasp. With the addition of more extensive subtests such as time based testing which "improves the test psychometric characteristics in order to detect dementia in the earliest stages" [9], we can increase our overall accuracy.

References

[1] I. Ognjanovic, "Artificial intelligence in healthcare." *Studies in health technology and informatics*, vol. 274, p. 196, 2020.

[2] V. Bloniecki, G. Hagman, M. Ryden, and M. Kivipelto, "Digital screening for cognitive impairment - a proof of concept study." *The journal of prevention of Alzheimer's disease*, vol. 8, no. 2, p. 129, 2021.

[3] C. Potts, J. Richardson, R. B. Bond, R. K. Price, M. D. Mulvenna, P. Zvolsky, M. Harvey, C. F. Hughes, and F. Duffy, "Reliability of addenbrooke's cognitive examination iii in differentiating between dementia, mild cognitive impairment and older adults who have not reported cognitive problems." *EUROPEAN JOURNAL OF AGEING*, 2021.

[4] K. L. Lanctôt, J. Amatniek, S. Ancoli-Israel, S. E. Arnold, C. Ballard, J. Cohen-Mansfield, Z. Ismail, C. Lyketsos, D. S. Miller, E. Musiek, R. S. Osorio, P. B. Rosenberg, A. Satlin, D. Steffens, P. Tariot, L. J. Bain, M. C. Carrillo, J. A. Hendrix, H. Jurgens, and B. Boot, "Neuropsychiatric signs and symptoms of alzheimer's disease: New treatment paradigms." *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 3, pp. 440 – 449, 2017.

[5] K. Wild, D. Howieson, F. Webbe, A. Seelye, and J. Kaye, "Status of computerized cognitive testing in aging: A systematic review," *Alzheimer's & Dementia*, vol. 4, no. 6, pp. 428–437, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S155252600802863X

[6] S. Müller, O. Preische, C. Laske, P. Heymann, and U. Elbing, "Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of alzheimer's disease from cognitively healthy individuals." *Frontiers in Aging Neuroscience*, vol. 9, no. APR, 2017.

[7] H. Takechi and H. Yoshino, "Usefulness of cogevo, a computerized cognitive assessment and training tool, for distinguishing patients with mild alzheimer's disease and mild cognitive impairment from cognitively normal older people." *Geriatrics —& Gerontology International*, p. 193, 2020.

[8] A. P. Charalambous, A. Pye, W. K. Yeung, I. Leroi, M. Neil, C. Thodi, and P. Dawes, "Tools for app- and web-based self-testing of cognitive impairment: Systematic search and evaluation." *Journal of Medical Internet Research*, vol. 22, no. 1, 2020.

[9] J. de Rotrou, L. Battal-Merlet, E. Wenisch, C. Chausson, E. Bizet, F. Dray, H. Lenoir, A.-S. Rigaud, and O. Hanon, "Relevance of 10-min delayed recall in dementia screening," *European journal of neurology*, vol. 14, no. 2, p. 144—149, February 2007. [Online]. Available: https://doi.org/10.1111/j.1468-1331.2006.01578.x