Behavioral Phenotyping for Autism Spectrum Disorder Biomarkers Using Computer Vision

James-Andrew R. Sarmiento and Prospero Naval Jr
University of the Philippines Diliman, Quezon City, Philippines
Email: jrsarmiento1@up.edu.ph, pcnaval@dcs.upd.edu.ph

***Abstract*—Analysis of the behavior of children is important for the early detection of developmental disorders such as Autism Spectrum Disorder (ASD) that is usually characterized by impairments in social and communication skills and repetitive and stereotyped behaviors as several studies about these behaviors revealed indicative of ASD in recorded videos of children later diagnosed with ASD. Developing a computational approach that is standardized and objective for the behavioral phenotyping of children with autism spectral disorder is significant for data gathering for creating prediction model for screening and monitoring of these patterns. We create a scalable, baseline application using computer vision algorithms and methodologies to capture four simple biomarkers data reliably for visual attention tracking using head pose estimation, specific observable behavioral patterns measuring blink rate and body posture, and morphological anomalies examining open mouth appearance from children with ASD. We conclude that it is feasible to quantitatively measure these behavioral phenotypes and there are promising results from measuring simple biomarkers showing distinguishing results for at least three biomarkers.**

***Index Terms*—computer vision, autism spectrum disorder, behavioral phenotyping, pattern analysis**

I. Introduction

Analysis of the behavior of children is important for the early detection of developmental disorders such as Autism Spectrum Disorder (ASD) as several studies about these behaviors revealed indicative of autism in recorded videos of children later diagnosed with autism [1]. ASD is a neurodevelopmental disorder that is usually characterized by impairments in social and communication skills and repetitive and stereotyped behaviors [2]. While the underlying cause of autism is still far from known, it is argued that many children with the disorder tend to exhibit several specific behavioral markers as early as infanthood and these symptoms can be observed during activities requiring visual attention where the child in response to a stimuli manifests difficulties in engagement and attention to these stimuli [1]. Children diagnosed with genetic syndromes such as autism disorder displays a characteristic and consistent cognitive, personality behavioral pattern that typifies a certain disorder [3]. Clinical observations are still the basis for these behavioral ratings for screening, diagnosis, and assessing neurodevelopmental disorders, including autism spectral disorder, which are highly subjective and require significant clinician expertise and training [4].

Developing a computational approach that is both standardized and objective for the assessment of behavioral markers of children with autism spectrum disorder is significant for data collection, data modelling, monitoring and screening of these patterns. Here, we discuss how computer vision can develop a scalable, low-cost application that could gather data from several biomarkers ranging from attention tracking, behavioral patterns, and morphological anomalies using a single video captured from an application in response to a visual stimulus.

The availability of scalable digital tools in measuring behavioral phenotypes will be vital in identifying and treating children with neurodevelopmental disorders such as in the case of ASD which affects one out of 59 children [4]. Early detection and diagnosis would allow early intervention that would substantially improve child behavioral outcomes and may have a greater impact preventing difficulty in different behaviors.

The main objective in this paper is to create a simple application to show the use of computer vision algorithms and methodologies to capture data reliably for visual attention tracking, specific observable behavioral patterns, and morphological anomalies from children with ASD.

II. Methods

There are numerous biomarkers for autism spectrum disorder. For this paper, we chose to present four biological markers where we capture data relevant to autism and other neurodevelopmental diseases. The video for analysis is captured while the child is watching a thirty-second movie clip as a visual stimulus from a stationary mobile phone on a landscape orientation.

A. Attention Tracking

Head pose is stated to be behavioral measurements related to attention in developmental disorders [5]. Differences in patterns of attention on a normal child and someone with developmental disorders are key behavioral indicator and could be collected as data. We track one behavioral movement identifying if the child is paying attention to a stimulus or not.

1) Head pose estimation

Manuscript received January 5, 2020; revised May 8, 2020.

©2020 Journal of Image and Graphics
doi: 10.18178/joig.8.2.47-52

It has been observed that head movement dynamics were greater in children without ASD although [6] findings show that the differences were evident in lateral movements (yaw and roll) but not vertical (pitch) movement. We collect head movement data using yaw, roll, and pitch. For each video frame, the head-pose estimation is predicted by the detection of 2D facial landmarks to estimate the 3D pose as seen in Image 1. We use dlib [7], a toolkit containing machine learning algorithms and tools for finding the face and 68 facial landmarks necessary for predicting the head pose using OpenCV. For each frame, we detect the 2D facial landmarks, identify image locations predict 3D pose, and gather roll, pitch, and yaw measurements as a time-series data.

Figure 1. Facial landmarks and head pose estimation with 3D box

Fig. 1 shows marked facial landmarks inside a 3D cube that gives the head pose estimate consisting of yaw, pitch, and roll.

B. *Tracking of Behavioral Patterns*

Certain behavioral patterns on people with autism have been observed such as increased eye blink rate [8], impaired motor abilities and asymmetric body positions [3], and repetitive motor movements (e.g. hand waving, body rocking hand flapping) [9]. Here we try to collect a time-series data for two specific biomarkers – blink rate and body pose estimation, using computer vision.

1) *Eye blink rate*

It is observed that there is an elevated blink rates on children with ASD suggesting increased dopaminergic activity in children [8]. For detecting eye blink in an image, we use the notion of Eye Aspect Ratio (EAR) where each eye is represented by six (x, y) coordinates as illustrated in Fig. 2.

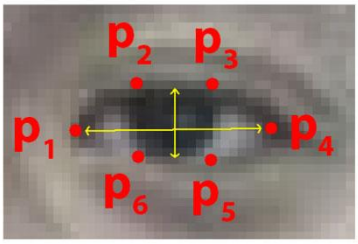

Figure 2. Six keypoints in the eye associated with EAR [10]

Using these coordinate points, we can derive the equation (1) that reflects the width and height of these coordinates as eye aspect ratio [11]. With this equation, we could predict that a person blinks when the eye aspect ratio decreases dramatically approaching to zero on a threshold. The EAR for a fully open eye is larger and relatively constant over time. Data is collected as a time-series data for the presence of eye blinks.

$$EAR = \frac{\| p_2 - p_6 \| + \| p_3 - p_5 \|}{2 \| p_1 - p_4 \|} \quad (1)$$

2) *Body pose estimation*

Body motion is recognized to be an important behavioral measurement for people with ASD which includes head turning, postural control, and asymmetric arm positions [3]. We could estimate 2D body pose and estimate arm and shoulder angles while the child is watching a visual stimulus. Observing patterns in body movements measured by angles in a time-series fashion in body movements measured by angles in a time-series fashion could provide insight for assessment and monitoring for motor movements of children with autism.

In order to measure these angles, the keypoints or landmarks of the body needed to be identified first. We use a pre-trained model created by [12] trained using the COCO Human Pose Dataset [13]. Cao *et al.* [12] created a real-time multi-person 2D pose estimation method by training a very deep Neural networks that takes a colored image of size w x h and produces the 2D location of keypoints for the person in the image. The model architecture imposes the following for producing 2D keypoints for people in the image.

- 10 layers of VGGNet to create feature maps for the input image
- A 2-branch of multi-stage CNN to predict a set of 2D confidence maps and affinity of body parts allocation (e.g. elbow, knee, etc.) and 2D vector fields encoding the association between parts (e.g. neck to left shoulder)
- A confidence and affinity maps are parsed by greedy inference

Using this model, we get the landmarks with multiple keypoints on frames on every specified time interval (e.g. every 0.1s). We aim to collect angles from left and right shoulders, and left and right elbows as illustrated in Fig. 3. The angles are calculated by first getting the slope of a line and finding the angle in degrees using the equation of the two adjacent lines. Each of the five angles are collected as a time-series data.

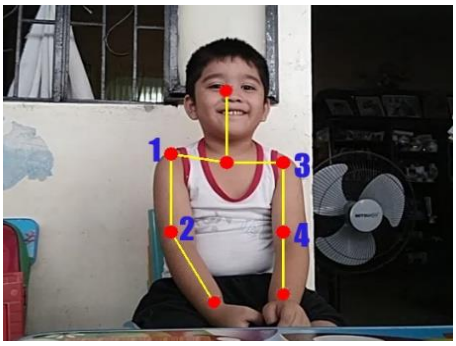

Figure 3. Body landmarks and angles measured (1) right shoulder, (2) right elbow, (3) left shoulder, (4) left elbow

©2020 Journal of Image and Graphics

C. Measuring Morphological Anomalies

Morphological abnormalities are seen to be significantly more prevalent in patients with autism than normal control group where a number of distinguishing morphological features are associated with autism which includes quantitative measurements of the body and face asymmetry, eyes, asymmetry, etc. [2].

1) Open mouth appearance

Several findings have demonstrated that there is a high frequency of morphological anomalies highly elevated on children with ASD which includes major and minor anomalies which can be direct or indirect [2]. Here, we examine the detection of open mouth expression as one indirect malformation for children with autism using the same technique as detection of Eye Blinking. We measure mouth aspect ratio (MAR) using the mouth landmarks located by the use of dlib and facial landmark detection. We collect the data for MAR for the whole time-series.

$$MAR = \frac{\| p_{51} - p_{59} \| + \| p_{52} - p_{58} \| + p_{53} - p_{57} \|}{2 \| p_{48} - p_{55} \|} \quad (2)$$

III. Results and Discussions

We use the video captured while each child is watching the same thirty second video clip as visual stimuli and feed it into the application. We measure head-pose estimation for attention tracking, half body pose estimation angles, blink rates for behavioral pattern and open mouth appearance for morphological anomaly.

For the context, the children are watching a sample video of a funny movie clip from Despicable Me showing the minions. We captured data from four people, two children affected with autism and two control. Also, we label children having ASD with A (e.g. A1, A2) and children without autism with B (e.g. B1 and B2). A1 tends to have a better socialization skill than A2 as socializing with A1 to be able to make him watch the video took drastically less time than A2. Moreover, while the target should have been children of younger age (i.e. less than five years old), the four children have ages ranging from 8 to 12.

A. Measuring Head-Pose Estimation

Head poses are measured every frame and plotted as a time-series data. As quick observation from Fig. 4, the pitch for both children with autism spectrum disorder (A1 and A2) are often not in neutral position, which is zero. A1 tends to look down as observed with a negative pitch while A2 tends to look up (high positive pitch) both consistent on their own head orientation all throughout the duration of the video. B1 who has no autism disorder tend to be the same as other normal child in the experiment where they tend to have a neutral head pose most of the time except on time when they are laughing in reaction to the video (e.g. between 25th to 30th second).

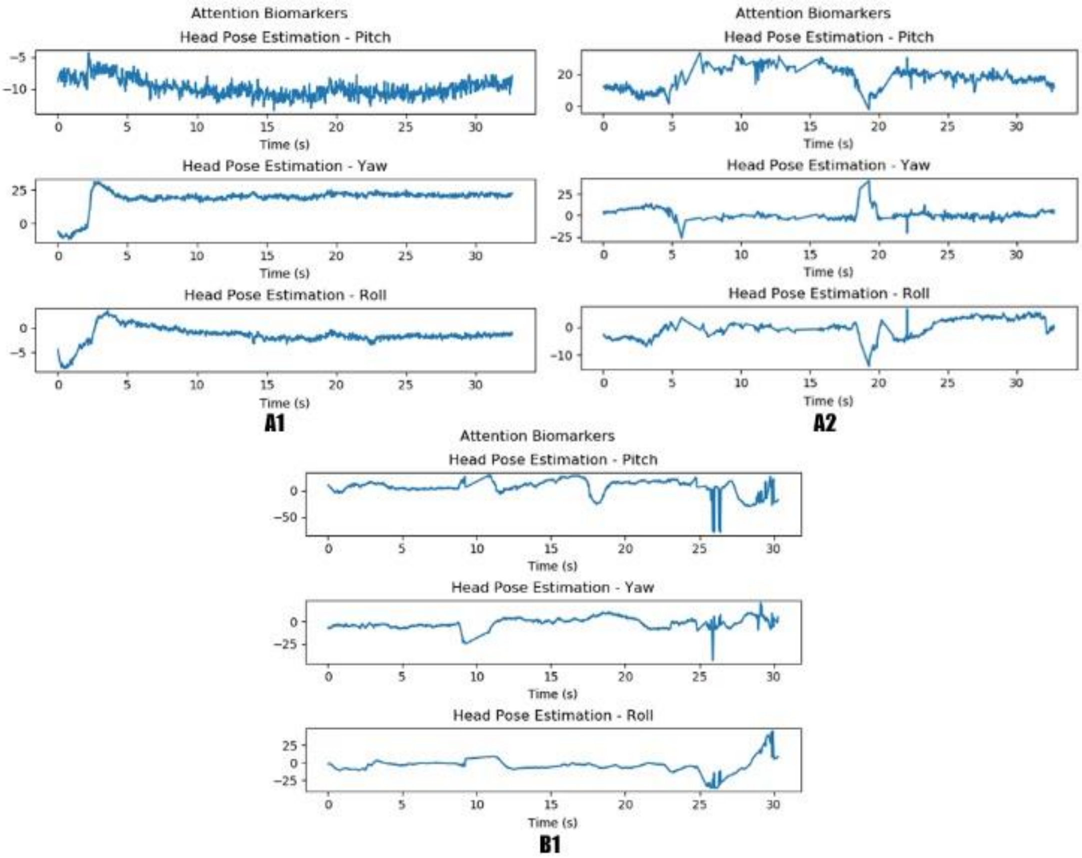

Figure 4. Pitch, yaw and roll time-series data for A1, A2, and B1

Another video is captured for A2 where Fig. 5 shows multiple, sudden, quick shifts of the head for both pitch, yaw and roll while she is trying to interact with the video.

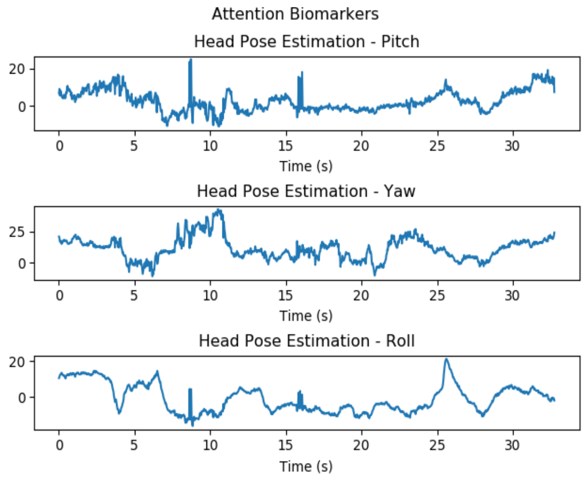

Figure 5. Second time-series data for A2

B. *Detecting Eye Blink Rate*

Here, we measure both EAR and Presence of eye blink over time. Using EAR has been a challenge for different participants. The typical EAR is around 0.30 for children with normal to wide eyes but is about less than 0.20 for those which has narrow eyes. Manual readjustment upon inspection of EAR throughout the time-series for the adjustment of blink threshold is done for participant A2 as she tends to have very narrow eyes. The tool works normally for A1 and the normal control group.

Interestingly, there is not much difference for the blink rates of each individual. A1 only had two recorded blinks while the other participants had blink count ranging from five to seven. We think that the video shown influences each reaction and that eye blink rate might not be a distinguishing feature for autism spectrum disorder. However, based on research [8], eye blink-rates could still be a factor so it is possible that there might be a pattern on this biomarker for a decent sample size and a more engaging visual stimulus.

C. *Examining Body Pose Angles for Asymmetry*

We try to examine four angles to check arm movement and arm asymmetry. The video has to be able to capture until the wrist to be able to detect the body landmarks. Meanwhile, the CNN is sometimes having a hard time detecting the forearm thus it creates wrong values occasionally on the elbow. Still, it detects all landmarks most of the time on a good background and the body parts are shown clearly on the video. Fig. 6 shows the body pose skeleton output for A2 generated.

We try to examine only data from A2 as all control groups have their measurement showing minimal movement on their arms while watching accompanied by symmetric arm positions. A2, most especially its second capture, displayed results indicating occasional movements of the arms and some asymmetrical angles.

Figure 6. Body pose skeleton output for A2

The arm position at the beginning is the cause for the asymmetric position of the arms. But all throughout, we see sudden shifts on these angles indicating arm and shoulder movements as seen in Fig. 7. We think that a proper starting posture is required for each participant and that there are video clips requiring them to do instructions. This way, the behavioral pattern might manifest. Unfortunately, we only observed the children just sit down thus it provides us data which has little to no difference for the bodily posture.

Figure 7. Time series data for A2 body pose estimation angles

D. *Observing Open Mouth Expression*

One indirect morphological anomaly observed upon children diagnosed with ASD is open mouth expression. We examine this by using MAR as discussed previously. In a quick glance, no pattern may be apparent. But a deeper look would show us that the mouth aspect ratio for A1, A2, and the second capture for A2 referred at Fig. 8 does not fluctuate drastically and are higher in average compared to B1 which only increases on sudden reactions to the video.

The MAR for A1 has a high value and does not vary from its mean by a very large margin typically only around 0.1. This may show an open mouth expression that does not provide any reaction indicating a change in facial expression. A2 on the other hand provides us drastic, varying changes on MAR on quick succession that may indicate sudden, multiple shifts on facial expression. B1 illustrates us some changes on MAR that indicates reaction to the video such as happiness and laughing with respect to certain parts of the video. Not only it could detect open mouth appearance as morphological anomaly, one could also derive change in facial expression using this approach to observe the response on a visual stimulus.

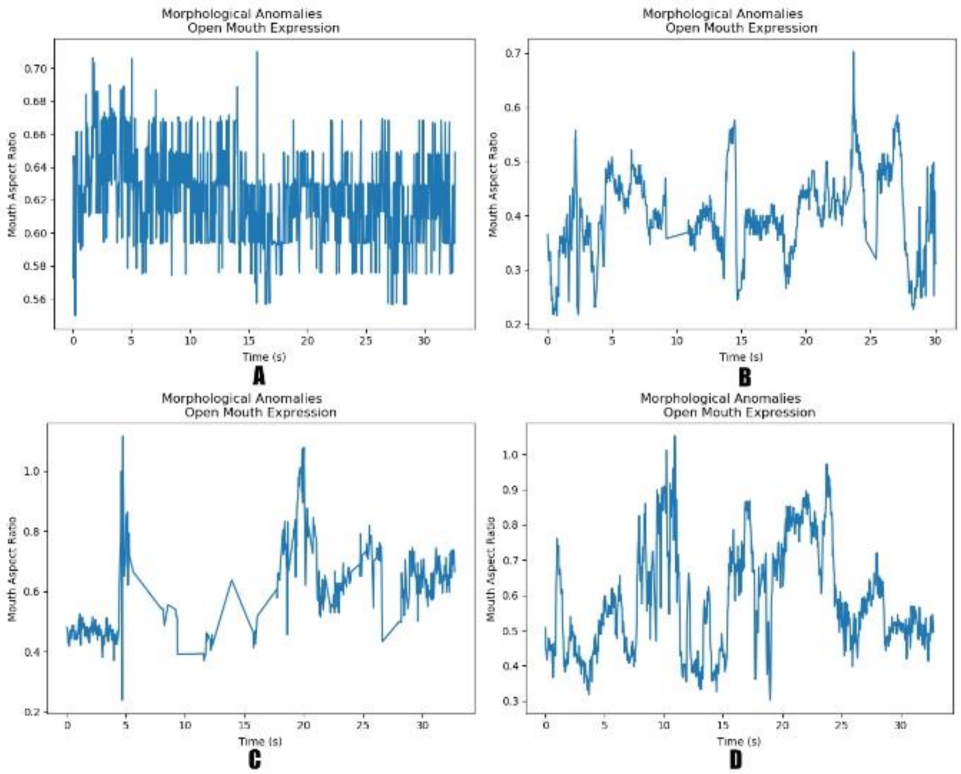

Figure 8. Mouth aspect ratio on time-series for (A), A1, (B) A2, (C) B1, and (D) second video on A2

IV. CONCLUSIONS

Autism Spectrum Disorder (ASD) is a neuro-developmental disorder where a child displays impairment on its social interaction skills accompanied with a characteristic repetitive behavior and morphological anomalies. We create a baseline tool which is scalable using computer vision to quantitatively measure into a time-series data some of basic distinguishing features which includes attention tracking using head pose estimation, behavioral patterns measuring blink rate and half body posture, and morphological anomaly examining open mouth appearance. Using a small sample size of children with and without autism, we observe that features like head-pose estimation, open mouth appearance and body posture estimation show promising differences from control group and one with autism spectrum disorder while eye blink rate does not give distinguishing differences. We suspect that these findings are influenced by a very small sample size and the nature of the visual stimuli the children are watching. Moreover, properly socialized child with autism still tend to show similar behavioral phenotypes as the control group on certain biomarkers such as body posture.

We conclude that it is feasible to use computer vision to measure observable characteristics of children which may have symptoms of autism spectrum disorder and that it is scalable enough to add modules in the future that focuses on other biomarkers. This tool could then be used for data collection for both creating the prediction model for screening and monitoring development of the child undergoing therapies. We think that the tool should work better for younger children but there should be innovation in ways of capturing the data in a standardized manner.

V. RECOMMENDATIONS

Since this is a collection of tools/program for getting data from biomarkers using a single video, it is encouraged to add other modules and tools for the collection of other behavioral patterns using computer vision such as emotion recognition, stimming movements such as hand flapping recognition. It is also recommended to look at non-visual clues such as speech recognition. We also recommend having variety of videos emphasizing focus on videos with multiple visual stimuli and videos promoting interaction to children to further observe difference in attention.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

James-Andrew Sarmiento conducted the research, analyzed the data, created the code. Both James-Andrew Sarmiento and Pros Naval Jr. wrote the paper. All authors had approved the final version.

ACKNOWLEDGMENT

The authors wish to thank Marek Ganko, for creating the mobile application to capture the data. Also the authors would like to thank UP Foundation for the funding support.

REFERENCES

[1] J. Hashemi, *et al.*, "Computer vision tools for low-cost noninvasive measurement of autism-related behaviors in infant," *Autism Research and Treatment*, vol. 2014, 2014.

[2] H. Ozgen, *et al.*, "Morphological featuers in children with autism specttrum disorders: A matched-case-control study," *Journal of Autism and Developmental Disorders*, vol. 41, no. 1, pp. 23-31, 2011.

[3] S. Cassidy and C. Morris, "Behavioral phenotypes in genetic syndromes: Genetic clues to human behavior," *Advances in Pediatrics*, vol. 49, pp. 59-86, 2002.

[4] J. Hashemi, G. Dawson, and G. Sapiro, "Computer vision and behavioral phenotyping: An autism case study," *Current Opinion in Biomedical Engineering*, vol. 9, pp. 14-20, 2019.

[5] K. Chawarska, S. Macari, and F. Shic, "Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders," *Biological Psychiatry*, vol. 74, no. 3, pp. 195-203, 2013.

[6] K. Martin, *et al.*, "Objective measurement of head movement differences in children with and without autism spectrum disorder," *Molecular Autism*, vol. 9, no. 1, p. 14, 2018.

[7] D. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755-1758, 2009.

[8] B. Jensen, *et al.*, "Increased eye-blink rate in autism spectrum disorder may reflect dopaminergic abnormalities," *International Society for Autism Research*, poster presentation, 2009.

[9] J. Matson, M. Gonzales, J. Wilkins, and T. River, "Reliability of the autism spectrum disorder-diagnostic for children (asd-dc)," *Research in Autism Spectrum Disorders*, vol. 2, no. 3, pp. 533-545, 2008.

[10] A. Rosebrock, "Six keypoint Landmarks associated with the eye," 2017.

[11] J. Chech and T. Soukupova, "Real-time eye blink detection using facial landmarks," *21st Comt. Vis Winter Work*, 2016.

[12] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291-7299.

[13] T. Lin, *et al.*, "Microsoft COCO: Common objects in context," in *Proc. European Conference on Computer Vision*, 2014, pp. 740-755.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

James-Andrew R. Sarmiento is an M.S. Computer Science student from the University of the Philippines. He is specializing in Computer Vision and Machine Learning in his graduate studies. He acquired his B.S. Computer Science degree from the University of the Philippines Manila.

He is currently a full stack developer at Code Ninja IT Solutions where he does full stack web development, data science, and machine learning projects. His interests include learning new technologies in Machine Learning and Deep Learning and he is also passionate in trying to learn the current technologies in web and app development.

Prospero C. Naval Jr. holds a Ph. D in Electrical from the University of the Philippines. He acquired his B.S. and M.S. degrees in Electrical Engineering from the same university. He also holds an M.Eng. in Information Science from Kyoto University, Japan.

He is currently a professor at the Department of Computer Science of the University of the Philippines, where he heads the Computer Vision and Machine Intelligence Lab. His research interests include underwater computer vision, data analytics for healthcare, environment, and education, deep reinforcement learning, deep learning, probabilistic machine learning, and swarm robotics.