An Eye-Tracking Solution Using Consumer Grade Webcams for Potential Concussion Diagnosis and Evaluation

Nicholas Caporusso*, Gabriel Sanders*, Brett Thaman*, Elaina Hall*
* Northern Kentucky University, Highland Heights, Kentucky, USA
{caporusson1,sandersg1,thamanb1,halle14}@nku.edu

***Abstract*—In the past decade, eye-tracking technology has been increasingly utilized in healthcare settings to assess an individual's cognitive status under various circumstances to evaluate the presence of brain injuries and monitor the evolution of temporary or permanent cognitive illnesses. More specifically, emerging research has confirmed that disconjugate eye movements can be utilized as a predictor of a concussion and potentially return-to-play protocols.**
Currently, most healthcare applications use eye-tracking devices requiring more detailed signals at a higher sampling rate and with better camera resolution compared to the eye-tracking technology used in the consumer market. Unfortunately, this limits the availability of this type of diagnostic system to clinical settings and prevents its use in situations where early diagnosis is crucial (e.g., during a football game, where 300,000 concussions are reported yearly).
In this paper, we introduce a solution for potentially diagnosing and treating concussions based on images acquired with inexpensive and more available camera devices such as webcams, and we detail a performance evaluation study of a popular image segmentation and object detection machine learning model (i.e., MediaPipe Facemesh and Iris) applied to the acquisition and analysis of eye-related signals (e.g., eye movements and blinking) for healthcare applications.

***Keywords*—component, formatting, style, styling, insert (key words)**

I. INTRODUCTION

A concussion is considered a mild traumatic brain injury that is caused by a bump or jolt to either the head or the body resulting in rapid brain movements inside the skull. Of the 1.4 million annual traumatic brain injuries (TBI) in the United States, 50,000 persons die and another 235,000 will require hospital admission while 50% of incidents go unreported or undetected [1]. More specifically, it is estimated that more than one-third of concussions go undiagnosed in athletes [2]. The lack of proper diagnosis of concussion may result in serious long-term consequences or risk of coma or death [3].

To avoid catastrophic injury and long-term negative health consequences due to undiagnosed concussions, accessible and affordable technology should be utilized and better developed for consumer and medical-grade consumption. Technology advances in the last decade have found eye-tracking can be used to assess brain injury and it can be utilized as a high-sensitivity biomarker compared to other types of subjective exams and assessments to determine concussion symptoms that are commonly administered by physicians [4] [5] [6].

Current eye-tracking solutions are expensive, are not accessible to many areas and people, and are only available via a healthcare provider. A frequent issue with non-computerized concussion assessments is the duration of the test and the likelihood of false-positive diagnoses or data to support a return to play or work [7]. For example, baseline testing batteries with reliable day-to-day change indices have merit, but invalid changes in a gait test or dexterity test can result in false-positive diagnoses due to faulty tests or interpretations [7] [8] [9]. Reliable and noticeable change indices can become even more challenging to interpret when implementing a multitest battery for concussion assessment [9]. There is emerging data and technology to support eye-tracking as a means for rapid diagnosis of concussions from youth to professional athletes [6] [10]. Eye-tracking can be a valuable tool to monitor subtle daily changes in eye movements that may indicate improvements in cognitive function resulting in well-calculated quantitative assessments for return to play, as opposed to subjective physician assessments. Neurocognitive testing is a method to assess brain function by implementing various noninvasive tests such as paper and pencil or computerized neurocognitive tests (CNT) [8] [11] [12]. More specifically, computerized neurocognitive testing can provide valuable concussion-related information only if a baseline test has been administered to serve as a reference for each individual [11]. In addition to a well-designed baseline test, these assessments can take a minimum of 15-30 minutes, need to be supervised by a clinical professional, and must be administered in an environment with no distractions with specific lighting that can also be administered on a sideline or in a locker room [13].

To avoid the complexity of cognitive assessments, rapid eye-tracking can be done more quickly, would not require a significant cognitive component, and could be used with minimal supervision as a computer program could guide the process. Rapid and reliable diagnosis of a concussion can occur on a field, in a gymnasium, or even at a residential home with portable and computerized technology designed to assess eye conjugation [5] [4] [14]. Accessibility to initial concussion diagnosis can further improve short and long-term health care by providing physicians with additional information regarding the extent of disconjugate eye movements [3] [10] [9] [15]. In addition to diagnosis, accessible eye-tracking technology can improve recovery assessments for patients as data can be collected daily to observe and record improvements in eye movements [10]. The information can then be relayed to a physician for a more accurate determination of return to play or return to work.

II. RELATED WORK

Several companies have developed commercial eye-tracking solutions based on expensive dedicated hardware that can be utilized to aid the diagnosis and treatment of concussions. The Oculogica EyeBox is approved by the Food and Drug Administration (FDA) to detect concussions, does not require a baseline assessment, and can be completed in 4-5 minutes [4] [5]. Additionally, due to its cost and acquisition processes, clinical-grade eye-tracking technology is only available at certain clinics or medical facilities, which limits the possibility of reaching people in rural areas and results in underserved populations.

Nevertheless, eye tracking is an established technology that utilizes sensors such as infrared or RGB cameras to track the pupil and its movement, it can be accomplished with consumer-grade web cameras. Unfortunately, given the cost, eye-tracking is not widely available, especially in high schools and underserved locations, where most traumatic brain injuries and concussions are undiagnosed. Most eye-tracking applications involve the use of dedicated external devices that range in price from \$250 to \$10,000 USD and can only be accessed via a healthcare provider. As previously mentioned, eye-tracking is used as a tool to diagnose concussions in a non-invasive, accessible, and quantitative way. Eye-tracking can also collect and obtain novel information about the severity of concussions that is unavailable from traditional, noncomputerized examinations alone [4]. When compared to the Sports Concussion Assessment tool version 3 (SCAT3) [15], eye-tracking produced results with both specificity and sensitivity that could be useful as a baseline-free means of diagnosis, as well as having the potential to be the standard for the detection of brain injury [5]. In patients with head injuries, eye-tracking revealed a correlation between the severity of the oculomotor disruption and the severity of the symptoms displayed [5]. Results showed that both the horizontal and vertical conjugacy of both eyes can be compromised due to a traumatic brain injury that was not present in non-trauma patients [5]. Applying eye-tracking to young athletes also yielded different results than commonly used clinical assessments, suggesting that it may be an objective addition to concussion examinations [10] [15]. With the correlation between eye movement and concussive symptoms, eye-tracking technology can be used to detect concussions with reasonable accuracy [4] [5] [6]. Studies have shown that there is a need for objective concussion examinations outside of a hospital environment, as many concussions go undiagnosed due to a lack of testing accessibility as well as errors in objective concussion examinations [2] [3] [14]. The abnormalities in the eye movements of concussed patients can be very subtle, which can make them difficult to subjectively detect even when diagnostic examinations are performed by trained professionals on cooperating patients [4] [5]. Common causes of misdiagnoses in patients examined by medical professionals include lack of transparency from patients, miscommunication between patients and professionals, and bias on the part of medical professionals and their equipment or the type of test [2] [4] [8] [15] [16]. Undiagnosed concussions can lead to health complications, such as an increased likelihood of comorbid events [17]. Eye-tracking can be used to prevent the consequences of undiagnosed concussions by making accurate concussion technology readily available to a wider range of patients and reducing qualitative errors.

III. EYE TRACKING USING MEDIAPIPE

In the last years, several groups explored viable alternatives to dedicated eye-tracking devices using traditional RGB cameras and image processing algorithms to support the wider use of eye-tracking technology in healthcare. A potential solution resides with recent developments in camera technology, suggesting that standard webcams can be utilized to realize accurate eye tracking at sampling rates and with resolutions that would be appropriate for healthcare applications. However, standard consumer cameras such as the webcams that are commonly incorporated in notebooks or used for video conferencing, have a lower acquisition frame rate (i.e., 24-30 frames per second), provide only a monocular view of the subject's face, and process the acquired images via software. Moreover, as described in [18], eye-tracking based on RGB cameras involves a sophisticated workflow consisting of multiple steps, that is, detecting and tracking the user's face, locating their eyes, identifying the position of their pupils, and subsequently estimating the coordinates of the point of the screen where the user is looking, in real-time. Indeed, each step involves a different problem, specific image processing techniques, and computational concerns and requirements, which significantly affect performance (i.e., speed), accuracy (i.e., the distance between the predicted gaze location and actual target), and reliability (i.e., within-subject and cross-subject accuracy).

More recently, MediaPipe [19] [20] has been introduced as an efficient solution for real-time computer vision tasks. It comprises a set of models optimized for specific image segmentation and object recognition problems such as face, hand, posture, and iris detection. Moreover, MediaPipe Iris [21] focuses on identifying a subject's pupils, tracking their movement, and estimating their distance from the camera.

MediaPipe FaceMesh predicts the facial topology of the user based on a still image and represents it as an array of 468 key points, that is, vertexes each associated with a set of three-dimensional coordinates that enable describing the components of the face (e.g., silhouette, nose tip, cheeks, and iris). Each property provides useful information about the positioning and alignment of the subject with respect to the camera, including face posture and rotation.

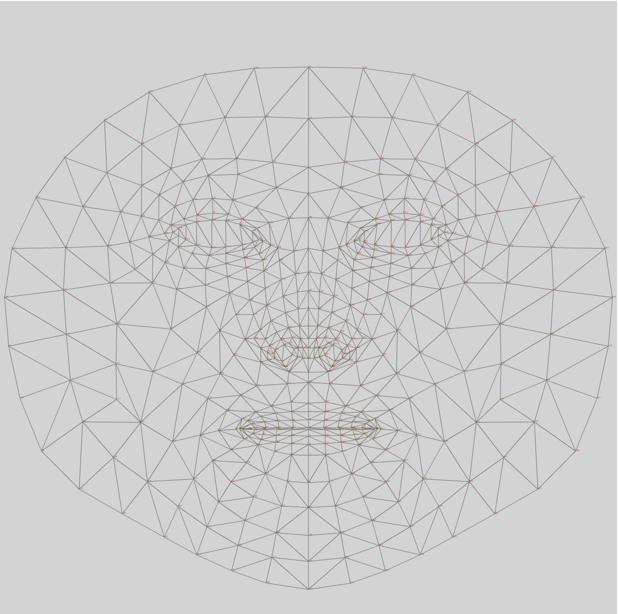

Fig. 1: The 478 landmarks utilized by MediaPipe FaceMesh and Iris to estimate the subject's facial geometry and the coordinates of their pupils.

In addition, MediaPipe Iris estimates the position of the user's pupils and represents them as an array of five landmarks for each eye (i.e., top, bottom, left, right, and center vertexes).

The opportunities offered by these models are remarkable, as MediaPipe FaceMesh and Iris are designed to be incorporated into cross-platform mobile applications and websites, which makes the library one of the promising candidates for exploring the feasibility of healthcare applications of eye-tracking based on standard RGB cameras. However, their clinical use has not been fully explored yet.

IV. Performance evaluation

The purpose of our work is to enable the development of healthcare applications based on the use of ML models such as MediaPipe FaceMesh and Iris. The ultimate goal of the proposed project is to develop a concussion assessment solution that would increase the affordability and availability of exams. If computer-grade webcams and Machine Learning models can predict eye features with sufficient accuracy, they could be advantageous in reassessing the subjects' conditions and managing their rehabilitation without expensive equipment. To this end, it is crucial to assess the model's performance in effectively detecting and tracking eye movements with a signal-to-noise ratio (SNR) that is appropriate for supporting diagnosing and monitoring cognitive conditions and brain injuries.

Therefore, we realized a performance evaluation study aimed at understanding the key factors that affect signal quality in terms of accuracy and reliability when using MediaPipe FaceMesh and Iris to estimate the facial geometry and position of the pupils from the images acquired with a standard webcam.

Fig. 2: One frame collected by the experimental software.

A. Materials and methods

In our data acquisition, we utilized five different notebooks incorporating 720p (1280 x 720 pixels) webcams with standard lenses. Also, we recorded data using an external 720p webcam mounting a wide-angle lens to compare the performance and evaluate any differences. We developed a dedicated data collection software that estimated the distance of the subject based on their facial landmarks by using the focal length of the camera. Also, the software recorded the location of the center of each pupil using a coordinate system relative to the size of the frame. Figure 3 shows a frame captured by the experimental software: the distance and frames-per-second are shown at the top left corner, whereas the bottom right corner shows an enhanced view of the right (red) and left (green) pupils. At the beginning of each trial, the software asked the subject to position themselves at a distance of 10 centimeters from the camera. Subsequently, every five seconds, the software asked the subject to move one distance step (i.e., 1 cm) away from the camera while keeping their eyes aligned with the lens. The trial stopped when the subject was at a distance of 110 cm from the camera. This is because a range of 10-110 cm is the typical distance at which individuals either stay when they use their computers or can be asked to maintain for the duration of a quick assessment (i.e., shorter end of the range). We did not use any physical guidance or constraint to keep the subject aligned with the camera's lens as they moved away from it. However, data recording was supervised to prevent the subject from moving too fast or significantly deviating from the ideal alignment with the camera. As each camera recorded at a rate of 30 frames per second, we collected 90-150 frames for each distance step from each subject, which we considered sufficient to evaluate SNR and signal dispersion.

Five healthy subjects participated in our data collection. A smaller pool of subjects was sufficient because the purpose of our study was to evaluate the performance of the system in terms of signal accuracy and reliability of webcam-based eye tracking. Each subject realized four trials with each camera, for a total of 20 trials per subject. As a result,

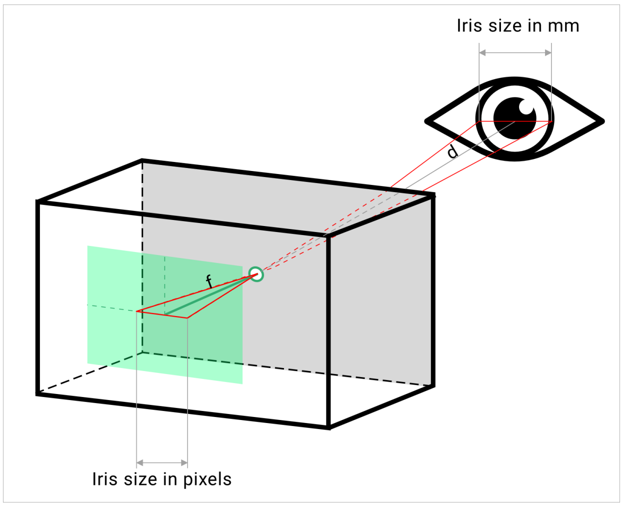

Fig. 3: Calculation of the distance of the subject based on the size of the iris and the focal length using similar triangles.

we acquired a total of 288,000 frames. After removing signal artifacts such as blinking, we had a total of 261,375 data points for each pupil (i.e., approximately 2600 frames for each distance step).

B. Results

Our data show that signal dispersion linearly increases with distance. The ideal SNR is achieved when the user is at a range of 10-50 centimeters from the webcam, where the dispersion is less than 4%, that is, less than 50 pixels on a 720p frame. Also, SNR degrades at a faster pace on the Y axis. This could be caused by vertical adjustments in the pose of subjects as they were moving away from the camera. Our findings are represented in Figures 4 and 5, which show the signal dispersion for the left and right eye, respectively.

Furthermore, by analyzing the variance of the signal dispersion, we can confirm that distances in the range of 10-50 centimeters are ideal for achieving more reliable SNR, with dispersion affecting the signal by less than 1 pixel, on average. Conversely, as shown by Figures 6 and 7, the signal shows wider dispersion when the subject is positioned more than 50 centimeters away from the camera. The dispersion variance, which is shown in pixels because of its small values, is contained within a range of 3.5 pixels on a 720pm webcam even at distances in the range of 100-110 centimeters, which shows the robustness of the facial geometry pipeline in accurately tracking the pupils.

Data from the wide-angle webcam show a statistically significant difference in the calculation of the distance of the subject from the camera, which is due to a different field of view (FOV) of the lens. However, our data did not show any statistically significant difference regarding SNR.

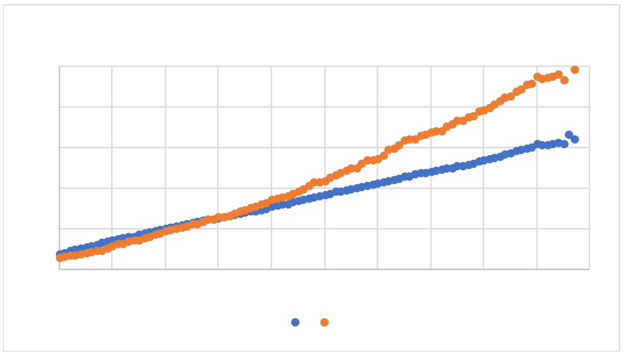

Fig. 4: Horizontal and vertical signal dispersion (percentage relative to the frame size) on the left eye.

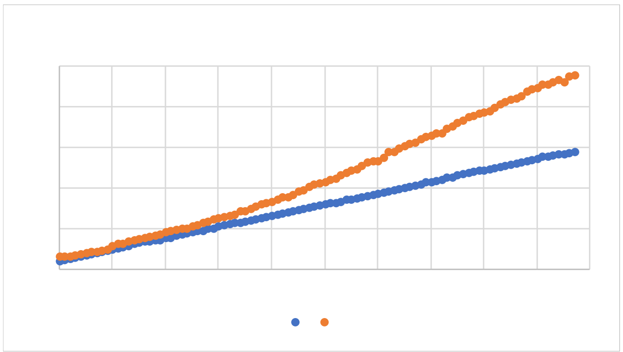

Fig. 5: Horizontal and vertical signal dispersion (percentage relative to the frame size) on the right eye.

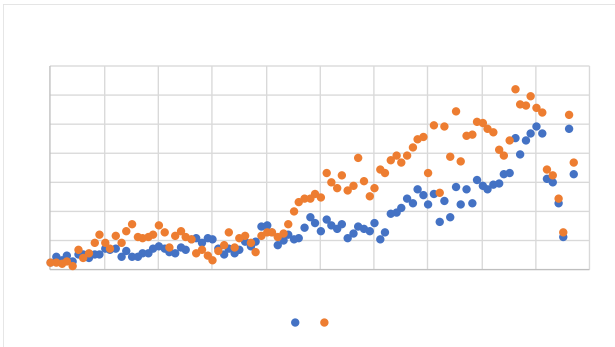

Fig. 6: Horizontal and vertical dispersion variance (in pixels) on the left eye.

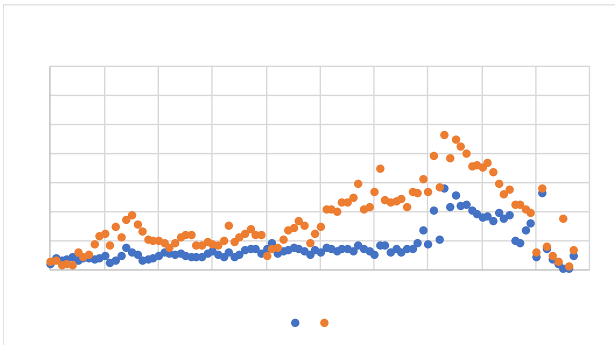

Fig. 7: Horizontal and vertical dispersion variance (in pixels) on the right eye.

C. Discussion

Our data show that MediaPipe FaceMesh and Iris are viable tools for potentially developing healthcare applications that leverage standard webcams for eye-tracking. Another contribution of our paper is that the subject's distance from the camera is a key factor. Based on our findings, the signal acquired by the system has sufficient accuracy and reliability when the subject is positioned at a distance ranging from 10 and 50 centimeters from the camera. Conversely, the SNR degrades significantly when the subject is more than 50 cm away from the camera, where noise values as high as 10% would not support the use of the system for the assessments of concussions. However, increased noise dispersion reported when the subject is more than 50 cm from the camera could also be caused by the subjects' movements as they were asked to distance themselves from the webcam.

Our experimental setup was designed to address the worst-case scenario, that is, circumstances in which subjects would not keep a fixed distance or hold their position in front of the camera. We were interested in evaluating the accuracy and reliability of MediaPipe FaceMesh and Iris at different distances from the camera and when the signal is affected by small movements of the subject to mimic real-world scenarios in which users would take their assessments at home, without supervision, or without adhering to recording protocols strictly. Even in this case, the system's performance supports tracking eye movements, including disjointed eye movements, fixations, and blinking, with sufficient accuracy. However, a higher precision assessment would require the subject to maintain their position at a specific distance (or within a defined range) from the camera, which can be obtained by implementing constraints based on the facial geometry estimated by MediaPipe FaceMesh. In addition to increasing the reliability of the signal acquired from the user, smoothing and denoising techniques could further enhance the signal. Moreover, we discovered that standard and wide-angle lenses can be used for data acquisition, provided that the focal length can be set as a configuration parameter to obtain more consistent results.

V. Conclusion

The goal of our work is to increase the availability of solutions for assessing and monitoring cognitive conditions and brain injuries such as concussions through the development of user-friendly technology that can be utilized more conveniently outside healthcare settings and without the direct supervision of clinical personnel.

In this paper, we presented a study in which we evaluated the performance of a Machine Learning model that could potentially be employed for estimating eye movements with a standard webcam. Specifically, we utilized MediaPipe FaceMesh and Iris, a cross-platform object segmentation and landmark detection library. Its advantage relies on leveraging standard webcams for estimating the facial geometry of the subject and tracking their eyes instead of requiring dedicated eye-tracking devices. However, as clinical assessments require high-quality signals, in our work we primarily focused on analyzing the accuracy and reliability of the data acquired by MediaPipe FaceMesh and Iris.

To this end, we realized a preliminary experiment in which we identified the ideal conditions for obtaining high-quality signals from MediaPipe FaceMesh and Iris. Our findings show that when the user is positioned at a range of 10-50 cm from the camera, the signal obtained by the webcam and processed by MediaPipe FaceMesh is sufficiently accurate and reliable for achieving a detailed assessment of eye movements, including disjointed eye movements, as discussed in previous studies. However, our data show that the SNR obtained by the system becomes very poor when the subject is located more than 50 cm away from the camera. Nevertheless, the information produced by the system is enough for adding constraints that provide the subject with guidance on the ideal positioning for signal collection.

In our future work, we will compare the performance of the proposed system with clinical-grade devices and realize control studies involving healthy subjects and individuals affected by brain injuries or cognitive conditions to assess the validity of our system and its potential uses in healthcare settings.

Acknowledgment

This project was realized with the financial support of NKU's Institute for Health Innovation and thanks to Kentucky Commercialization Ventures' IMPACT grant.

References

[1] L. M. Gessel, S. K. Fields, C. L. Collins, R. W. Dick, and R. D. Comstock, "Concussions among united states high school and collegiate athletes," *Journal of athletic training*, vol. 42, no. 4, p. 495, 2007.

[2] W. P. Meehan III, R. C. Mannix, M. J. O'Brien, and M. W. Collins, "The prevalence of undiagnosed concussions in athletes," *Clinical journal of sport medicine: official journal of the Canadian Academy of Sport Medicine*, vol. 23, no. 5, p. 339, 2013.

[3] M. Putukian, "The acute symptoms of sport-related concussion: diagnosis and on-field management," *Clinics in sports medicine*, vol. 30, no. 1, pp. 49–61, 2011.

[4] U. Samadani, M. Li, M. Qian, E. Laska, R. Ritlop, R. Kolecki, M. Reyes, L. Altomare, J. Y. Sone, A. Adem *et al.*, "Sensitivity and specificity of an eye movement tracking-based biomarker for concussion," *Concussion*, vol. 1, no. 1, 2016.

[5] U. Samadani, R. Ritlop, M. Reyes, E. Nehrbass, M. Li, E. Lamm, J. Schneider, D. Shimunov, M. Sava, R. Kolecki *et al.*, "Eye tracking detects disconjugate eye movements associated with structural traumatic brain injury and concussion," *Journal of neurotrauma*, vol. 32, no. 8, pp. 548–556, 2015.

[6] A. B. Zahid, M. E. Hubbard, J. Lockyer, O. Podolak, V. M. Dammavalam, M. Grady, M. Nance, M. Scheiman, U. Samadani, and C. L. Master, "Eye tracking as a biomarker for concussion in children," *Clinical journal of sport medicine*, vol. 30, no. 5, pp. 433–443, 2020.

[7] L. D. Nelson, "False-positive rates of reliable change indices for concussion test batteries: a monte carlo simulation," *Journal of Athletic Training*, vol. 50, no. 12, pp. 1319–1322, 2015.

[8] J. Bogner, L. Brenner, B. Kurowski, J. Malec, D. R. Howell, D. L. Hunt, J. R. Oldham, S. E. Aaron, W. P. Meehan, and C. O. Tan, "Postconcussion exercise volume associations with depression, anxiety, and dizziness symptoms, and postural stability: preliminary

findings," *Journal of head trauma rehabilitation*, vol. 37, no. 4, pp. 249–257, 2022.

[9] J. R. Oldham, D. R. Howell, C. A. Knight, J. R. Crenshaw, and T. A. Buckley, "Single-task and dual-task tandem gait performance across clinical concussion milestones in collegiate student-athletes," *Clinical journal of sport medicine*, vol. 31, no. 6, pp. e392–e397, 2021.

[10] J. R. Oldham, C. L. Master, G. A. Walker, W. P. Meehan III, and D. R. Howell, "The association between baseline eye tracking performance and common concussion assessments in high school football players," *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 98, no. 7, p. 826, 2021.

[11] W. P. Meehan III, P. d'Hemecourt, C. L. Collins, A. M. Taylor, and R. D. Comstock, "Computerized neurocognitive testing for the management of sport-related concussions," *Pediatrics*, vol. 129, no. 1, pp. 38–44, 2012.

[12] J. E. Resch, M. W. Schneider, and C. M. Cullum, "The test-retest reliability of three computerized neurocognitive tests used in the assessment of sport concussion," *International journal of psychophysiology*, vol. 132, pp. 31–38, 2018.

[13] R. C. Lynall, K. G. Laudner, J. P. Mihalik, and J. M. Stanek, "Concussion-assessment and-management techniques used by athletic trainers," *Journal of athletic training*, vol. 48, no. 6, pp. 844–850, 2013.

[14] M. Putukian, "Clinical evaluation of the concussed athlete: a view from the sideline," *Journal of athletic training*, vol. 52, no. 3, pp. 236–244, 2017.

[15] J. R. Oldham, W. P. Meehan III, and D. R. Howell, "Impaired eye tracking is associated with symptom severity but not dynamic postural control in adolescents following concussion," *Journal of sport and health science*, vol. 10, no. 2, pp. 138–144, 2021.

[16] S. V. Ettari, E. Roden, V. Ahuja, and U. Samadani, "Oculogica: An eye-catching innovation in health care and the privacy implications of artificial intelligence and machine learning in diagnostics for the human brain," *SMU Sci. & Tech. L. Rev.*, vol. 25, p. 23, 2022.

[17] T. McAllister and M. McCrea, "Long-term cognitive and neuropsychiatric consequences of repetitive concussion and head-impact exposure," *Journal of athletic training*, vol. 52, no. 3, pp. 309–317, 2017.

[18] M. S. Mounica, M. Manvita, C. Jyotsna, and J. Amudha, "Low cost eye gaze tracker using web camera," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, pp. 79–85.

[19] I. Grishchenko and V. Bazarevsky, "Mediapipe holistic—simultaneous face, hand and pose prediction, on device," *Retrieved June*, vol. 15, p. 2021, 2020.

[20] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *arXiv preprint arXiv:1907.06724*, 2019.

[21] A. Ablavatski, A. Vakunov, I. Grishchenko, K. Raveendran, and M. Zhdanovich, "Real-time pupil tracking from monocular video for digital puppetry," *arXiv preprint arXiv:2006.11341*, 2020.