International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-12 Issue-5, April 2023

A Convolutional Neural Network Study on Depression and Eye Blink Analysis

Bryan G. Dadiz, Alfio I. Regla

Abstract*: *This study is determining the correlation of human blinks relating to depression. The study uses convolutional neural network for detecting blinks in a video. Using Closed Eyes in the Wild dataset the Convolution Neural Network model was trained having 99.24% in training accuracy and 0.0275 loss from epoch of 50. However, the results from validation of the model resulted 61.09% tested from two datasets that where labelled with BDI-II depression scale. The study collated the results of recorded blinks from the video datasets and it showed that there is a weak positive correlation of the recorded blinks computed as blinking rates to depression. The result showed that the r2 score was 0>3.4 thus, there is a possibility but not the highly indicator of depression.

Keywords*: *AVEC' 14 Introduction, Convolution Neural Network, Depression Analysis, Eye Blinks, Major Depressive Disorder, Non-Verbal Behavior, Visual Cue

I. INTRODUCTION

Depression is also known as Major Depressive Disorder or Clinical prevalent problem in our modern day. It is a one of those global public health concern and it is directly associated with correlatively high connection to permanent disability [1-2]. It is stated by the World Health Organization that MDD is the 4th leading cause of disability worldwide and projects that by 2020 it will be the second leading cause of early mortality due to recent inexplicable vast prevalence in current related events [3-4]. A Lack of treatment for this mental health issue can be damaging and overtime be associated with serious physical disorder [5-6]. Moreover, Depression can cause major economic losses because depression was found to be correlated with decreasing productivity in working individuals, low educational attainment in students, high risk of teen pregnancies, break-ups among couple and poverty because of unemployment [7]. In United States, MDD directly increased burden in the economy, working individuals have increased their economic burden by 21.5% (from USD173.2 billion to USD210.5 billion) The cost remained stable with approximately 45% with attributes such as 5% to suicidal-related costs, and 50% to workplace costs. There is 38% of the total cost is directed to MDD itself in contrast to comorbid conditions. Therefore it correlates to the overall economic performance of a country [8]. On the other hand, psychological treatments of MDD has been proven effective [9], most of the time misdiagnosis is still a common barrier, as of the present practice [10]. Due to a lack of resources and trained health care providers this is the direct cause of the said barriers mentioned above, reported by the WHO. It is observed that there is a subjective bias on the present practice whereas the patients are self-reporting their symptom severity using assessment tasks or forms. The goal of this study is to investigate a nonverbal behavior of a person which will to an objective sensing hints of depression [11]. The results will help increment multimodal systems in identifying MDD to assist clinicians in their diagnosis and for monitoring MDD positive patients. In addition, the investigation of the non-verbal action such as the blink rate of clinically annotated patients with corresponding depression scale. Thus the proponents proposed a method whereas OpenCV library is utilized for extracting face and segmenting the eye region. Furthermore, the research examines the performance of these parameters using the convolutional neural network trained with Closed Eyes in the Wild (dataset) in detecting eye blinks within the provided video dataset of Audio Visual Emotion Challenge repository [12].

II. BACKGROUND OF THE STUDY

The study of understanding human behavior primarily in depression depends on observable self-report measures. Recent computational approaches provide possibilities for measuring computer-based audiovisual calculation of behaviors that human having difficulty to quantify (e.g. emotion, body movements and speech audial patterns). Having these tools, it can lead to potentially help clinicians and improve measure response and also diagnosis and overall screening process. Moreover, new patterns and subtleties could be tested in analyzing behavioral indicators in depression [13]. There have been researches that focused on non-verbal signals of depression from visual cues. facial expressions analysis from visual perspective [14], posture, body movements, head estimation, and gesture are visually communicated [15][30], Regardless, there are further developments currently in line, the studies that were presented recently have correlated on the same fundamental structures of analysis. Initially, the region of interest for body parts (e.g., head, torso, or hands) are being identified within each video frame. This procedure is usually reached by detecting within the video frame the localization of pre-trained models of certain parts of the body.

Manuscript received on 04 March 2023 | Revised Manuscript received on 16 March 2023 | Manuscript Accepted on 15 April 2023 | Manuscript published on 30 April 2023.
*Correspondence Author(s)
Bryan G. Dadiz*, College of Computer Studies, Technological Institute of the Philippines, Manila, Philippines. E-mail: bryan.dadiz@tip.edu.ph, ORCID ID: https://orcid.org/0000-0002-6802-5955
Alfio I. Regla, College of Computer Studies, Technological Institute of the Philippines, Manila, Philippines. E-mail: aregla.it@tip.edu.ph

Then, attributes are obtained from these measured and located parts that gives a number for their shape and/or form. Moreover, to lessen the noise there is a need to implement orientation-sensitive filters to the extracted body part to remove similar to primate vision errors [16]. In addition, a common practice is to manipulate the head estimation or pose size, attributes are typically registered to a common view then, Finally, an algorithm is established to interpret the features. Normally supervised machine learning is usually being used to achieve this phase, wherein samples of pre-captured or pre-identified images of a person is provided to a process of regression or classification. This will learn pattern in between features with a more diverse behavioral proportions or classes; Inference from this learned mapping can be used to a novel video in its frames. The latest findings say that, when used to discrete facial movements, These techniques are effective to variations in participant with ethnical background, same as the series of head positions, estimation and illumination varies common in abrupt data [17]. Most eye analyzed findings for eye movements are: (1) saccades, which are the fast movements of eyes while looking at a material (e.g. reading); (2) eye movements while looking at a moving object, (3) while sleeping eyes are examined, rapid eye movements (R.E.M.). The movement of the eyes are typically quantified with an electrooculography, a couple of electrodes mounted to the subject's eyes, then produces an electrooculogram signal [18]. A study found out that a horizontal pursuit movement of the eyes in depressed patients were founded abnormal from healthy controls [19]. It was confirmed by the study of Abel [20], states that the eye movements such as saccade and pursuit rates correlate strongly in controls but are reduced or lacking in subjects with emotional disorder, concluding abnormality in subjects' ocular motor systems as a sign of cognitive and physical movement retardation. Followed up by another study that abnormal saccades in patients were found even not undergone to any medication. Additionally, the outcomes in [21] identify both eye motor system disturbances and cognitive performance. R.E.M was analyzed consequently finding that latencies are reduced in people with depression and suggests that altered R.E.M latency can be regarded as measure for the severity of a mental illness [22]. Eyes when blinking was investigated that shows elevated blinks per second are present in depressed subjects, which becomes normal as the condition of the patient normalizes [23]. The area of machine learning specifically its branch called computer vision has a vast offering of variety techniques to be used for eye movement measurements such as camera markers for the eye, a mounted device for the corneal reflex illumination or the contact lenses method etc. [24]. This study is primarily focused in non-obtrusive procedures. Thus, the analysis of eyes movement in a video, initially it needs to find the head then find the eye with feature points to be used as features. The usual case is to detect the contour between the sclera and iris also called as limbus [25]. Furthermore, semi-automatic methods such as Active Appearance Models a generative method an example of a Fuzzy based approach for representing deformable visual objects could be trained to classify and localize human eyes [26]. The approach called fuzzy based the extracted AAM, from the eye variations such as half open, open and closed eyes needs to be tagged with a label before the AAM can be put into the training model. The trained model will then be implemented to search or identiy the postion of the eyes from the video. In this paper, the study used trained CNN to model and track the eye blinks with Haar segmentation techniques for localizing the eyes before classification in order to compute the variations between melancholy patients and the controlled samples of people. The eye movement movements of the sample videos in a recorded subject is doing human computer interaction tasks. Beside analyzing the general variations in eye blinks, study specifically investigate the performance of the CNN based model to test its accuracy from the Audio and Visual Emotion Challenge dataset comparing it to fuzzy based systems performance.

III. METHODOLOGY

A. The AVEC Dataset

The proposed study is computed using the Audio/Visual Emotion Challenge (AVEC) 2014 dataset [13] the dataset to be used is part of its subcomponent for the audio-visual depressive language corpus. The set of data is composed of a number of clips from 292 subjects. Using a webcam and a microphone the 340 video clips are people subjects that are performing Human-Computer Interaction tasks which were PowerPoint guided. A person is recorded in each video clip and sometimes have one or more number of sampled records. Wholly the subjects are recorded between one and four times, with twice a week interval. The overall mean 31.5 for their age with a standard deviation of 12.3 as their age coming from the range of 18 to 63 of age. The AVEC 2014 also is consisting of double tasks which are 14 in total from the original recordings. Thus, there are two separate footages in total of three hundred (300) videos with the videos scales from six seconds to four minutes and eight seconds. It is composed of two selected tasks based on maximum conformity. There are two samples which is Freeform those participants are responding on questions from the PowerPoint guided tasks while being recorded and Northwind which people who are reading aloud from a fable.

B. The Validation of the dataset

As for the labels of depression the videos from Freeform and Northwind have been pre-labelled by the use of a self-assessed subjective depression questionnaire, the Beck Depression Inventory-II (BDI-II) [29]. The form contains 21 questions, whereas in each question having a discrete scale with values from 0 – 3. The recordings of the concluding scores ranges from 0-63. The scale of range were interpreted to be 0-13: that there is no minimal depression, now 14-19: is mild depression, 20-28: shows moderate depression and 29-63: is an indication of severe occurrences of depression.

C. Convolution Neural Network for Detecting Blinks in a Video

Network for detecting open and closed eye adding up a threshold to counter eyeblinks. The CNN model is trained using the CEW dataset [12]

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

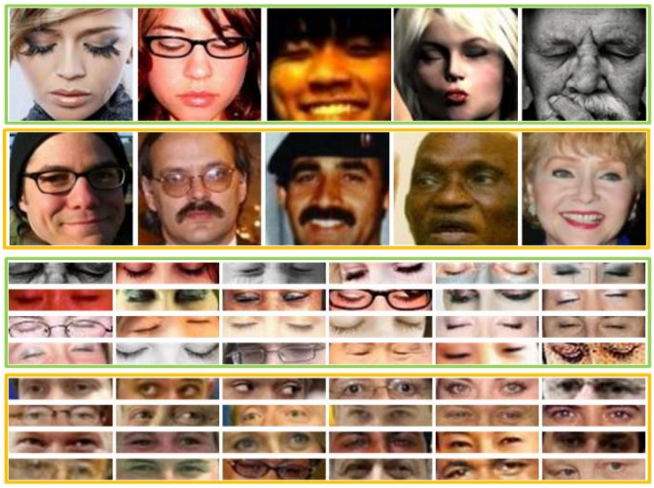

Fig 1. Illustration of the images in CEW dataset

In the above Figure 1 are the illustrations of CEW dataset. The top row are the faces with both eyes closed, while the row below it shows faces with eyes open. At the bottom of the second row are the sampled eye patches. Note that these sampled eye patches are having full variance of lighting, blur, occlusions and individual.

D. Convolution Neural Network for Detecting Blinks in a Video

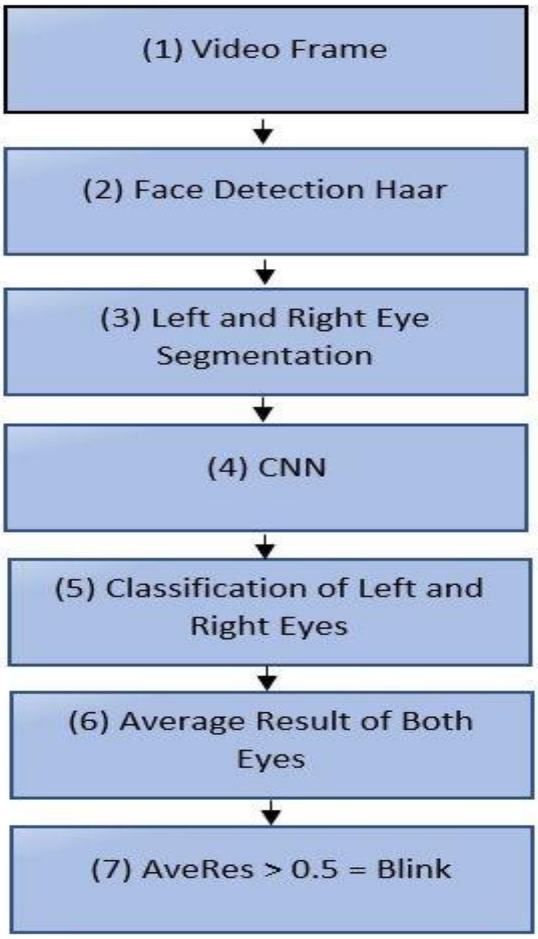

Fig 2. Proposed Approach for Blinks Detection for Depressed Subjects

Shown in Figure 2 the overall proposed approach presented in a sequence diagram or flowchart for the detection of the eye blinks while the eyes close and opens. Initially, from a video a frame is selection based on Haar classifier is used to detect the face and then right and left eye. The detected right and left will become the input image and cropped to a standard 26/34-pixel image. The left and the right segmented image from the previous step is used as the testing input for CNN classification of eye blinks. A more detailed discussion of the modelled CNN operation shall be explained in Section E.

E. Feature Extraction via Convolution Layers

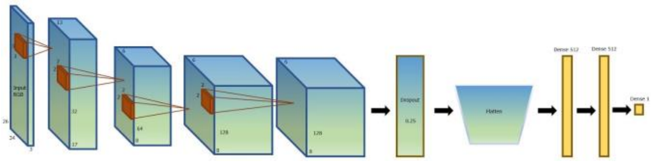

Fig 3. Convolution Neural Network Model

As shown in Figure 3, the first convolution in this model was inputted an image dimension of 26×34×3 and utilizes 32 filters that explores the vertical and horizontal directions of the inputted image meanwhile it strides by a two unit of pixel, having three-pixel units of padding in the vertical and horizontal directions. The second convolution performs operation with 64 filters having a size of 6x8x64 using a filter of 2x2 meanwhile it strides by a two unit of pixel. Lastly convolution three performs operation with 128 filters of 6x8x128 and traverse in the convolved image vertical and horizontal direction meanwhile it strides by a two unit of pixel. After each convolution a max pool is performed whose filter is 2x2 with a stride of two-pixel unit. Therefore, the feature map concludes to 1×1×512, which is implemented as the input of the fully connected layer.

F. Classification Layers

The total elements in the fully connected layers are linked to values of all units in its connected layers. For each link is needed for the calculation the needed weighted sum. The classification standard implemented the certainty from the weighted sum. In the layers when it reaches 1 as the final sum it increases the certainty of the detected image.

IV. RESULTS FROM THE EXPERIMENT

A. Accuracy of the Model

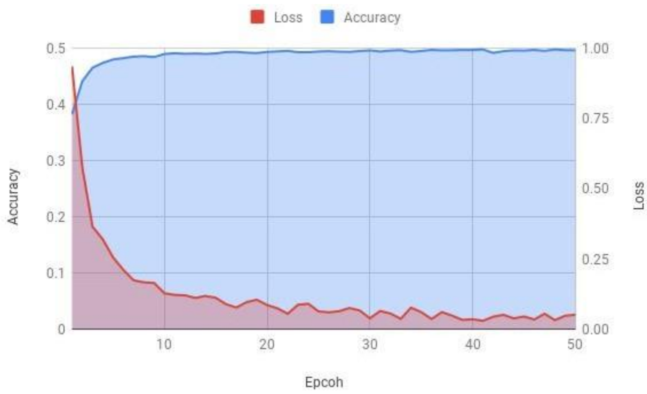

Fig 4. Graph of Loss and Accuracy for each Epoch

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Seen on Figure 4 the accuracy and loss on each epoch during the process of training by the model on Figure 3. For all the cases, the loss values reaches 0 while the accuracy of the training approaches 98% when the epochs of the training increases. It is being demonstrated here that the CNN model is sufficient implementing the CEW dataset as training data. The final loss is now 0.0275 while the accuracy of the training's final value reached 99.24%

B. Testing the Model

Fig 5. Testing the CNN Based Algorithm on Recorded Videos from Freeform and Northwind corpus.

As shown on Figure 5 the dataset from the Freeform and Northwind dataset whereas the CNN based detection of closed or open eyes is utilized in counting for human blinks from the total duration of recorded videos. Challenges were encountered whereas in the implementation of the algorithm have not properly detected blinks. In the experiment there were encountered some misclassification of blinks from the given dataset whenever the subject is moving rapidly with or without making eye blinks [27][28]. In the experiment, the videos that results were not properly classified were eliminated and computed as the validated error of the model. 52.09% from the total Freeform dataset were classified correctly while in the Northwind dataset it 70.9 were accurately classified.

C. Results of the correlation of Blinking from BDI-II Depression Scale

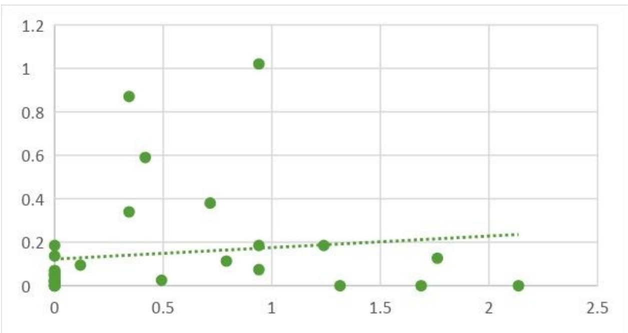

Fig 6. Northwind Correlation of Blinks to Depression

Shown in Figure 6 the result of Blink Rate and the labelled depression scale of the recorded subjects. The graph shows a weak correlation of blink rates from labelled depressive people as data was extracted from a CNN based blink recognition method on the Northwind dataset. The R2 value correlating blink rates from labelled depressed people resulted to 1%.

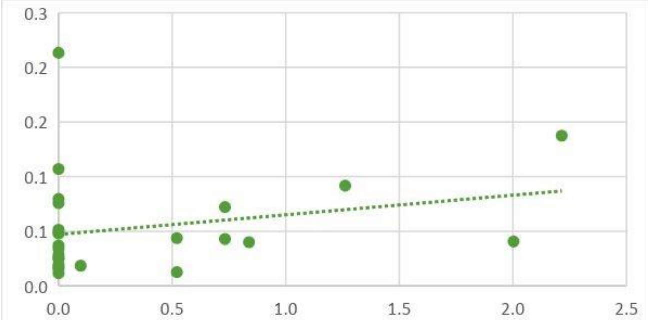

Fig 7. Freeform Correlation of Blinks to Depression

Shown in Figure 7 the result of Blink Rate and the labelled depression scale of the recorded subjects. The graph shows a weak correlation of blink rates from labelled depressive people as data was extracted from a CNN based blink recognition method. The R2 value correlating blink rates from labelled depressed people.

V. CONCLUSION

Based on the above results the implementation of the CNN model in detecting blinks from videos that were labelled as depressed resulted to 61.09% combined Northwind and Freeform. On the other hand, the datasets as shown from the results of two correlation graphs Figure 6 and Figure 7 had a weak correlation result combined Northwind and Freeform results having a weak positive correlation of r2 = 3.5. Therefore, the study shows that blink rate has weak positive correlation from these given datasets however there is still possibilities that by having more data might increase the correlation because of its hints of having positive correlation. Moreover, the application of CNN as a detector of eye blinks were found to perform well but it is recommended to apply deeper models of CNN and an increase of dataset for absolute detection of eyes while moving as an identified misclassification of the CNN model.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive it.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Ustun TB, Chatterji S. Global burden of depressive disorders and future projections. In: Dawson A, Tylee A, eds. Depression: social and economic timebomb. London: BMJ Books, 2001.

Retrieval Number: 100.1/ijitee.E94880412523
DOI: 10.35940/ijitee.E9488.0412523
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

2. Ustun TB, Ayuso–Mateos JL, Chatterji S, Mathers C, Murray CJ. Global burden of depressive disorders in the year 2000. Br J Psychiatry 2004; 184: 386–92. [CrossRef]
3. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
4. Murray CJL, Lopez AD. The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020.Harvard University Press; 1996
5. Solomon DA, Keller MB, Leon AC, et al. Multiple recurrences of major depressive disorder. Am J Psychiatry 2000; 157: 229–33. 5 [CrossRef]
6. Andrews G. Should depression be managed as a chronic disease? BMJ 2001; 322: 419–21. [CrossRef]
7. Ronald C. Kessler PhD Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA
8. Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). The Journal of clinical psychiatry, 76(2), 155-162. [CrossRef]
9. L. G. Kiloh, G. Andrews, and M. Neilson, “The long-term out come of depressive illness,” The British Journal of Psychiatry, vol. 153, pp. 752–757, Dec. 1988. [CrossRef]
10. N. Niedermaier, E. Bohrer, K. Schulte, P. Schlattmann, and I. Heuser, “Misdiagnosed patients with bipolar disorder: comorbidities, treatment patterns, and direct treatment costs,” The Journal of clinical psychiatry, vol. 66, no. 11, pp. 1619– 1623, 2004. [CrossRef]
11. J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, “Voice acoustic measures of depression severity andtreatmentresponsecollectedviainteractivevoiceresponse technology,” Journal of Neurolinguistics, vol. 20, no. 1, pp. 50–64, 2007. [CrossRef]
12. F. Song, X.Tan, X.Liu and S.Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, Pattern Recognition, 2014. [CrossRef]
13. M. F. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. AVEC 2014 [CrossRef]
14. 3d dimensional affect and depression recognition challenge. In International Conference on ACM Multimedia - Audio/Visual Emotion Challenge and Workshop, 2014.
15. Dadiz, B. G., & Ruiz, C. R. (2019). Detecting Depression in Videos using Uniformed Local Binary Pattern on Facial Features. In Computational Science and Technology (pp. 413-422). Springer, Singapore. [CrossRef]
16. Cummins, N., Joshi, J., Dhall, A., Sethu, V., Goecke, R., & Epps, J. (2013, October). Diagnosis of depression by behavioural signals: a multimodal approach. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge (pp. 11-20). ACM. [CrossRef]
17. Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T: A theory of object recognition: Computations and circuits in the feed forward path of the ventral stream in primate visual cortex. Artificial Intelligence 2005:1–130.
18. J.M. Girard, J.F. Cohn, M.A. Sayette, L.A. Jeni, F. De la Torre : Spontaneous facial expression in unscripted social interactions can be measured automatically. Behavior Research Methods In Press.
19. N. Kathmann, A. Hochrein, R. Uwer, and B. Bondy, “De□cits in gain of smooth pursuit eye movements in schizophrenia and affective disorder patients and their unaffected relatives,” The American Journal of Psychiatry, vol. 160, no. 4, pp. 696–702, 2003. [CrossRef]
20. R. B. Lipton, S. Levin, and P. S. Holzman, “Horizontal and vertical pursuit eye movements, the oculocephalic re□ex, and the functional psychoses,” Psychiatry Research, vol. 3, no. 2, pp. 193–203, 1980. [CrossRef]
21. L.A. Abel, L. Friedman, J. Jesberger, A. Malki, and H.Y. Meltzer, “Quantitative assessment of smooth pursuit gain and catch-up saccades in schizophrenia and affective disorders,” Biological Psychiatry, vol. 29, no. 11, pp. 1063–1072, 1991. [CrossRef]
22. T. J. Crawford, B. Haeger, C. Kennard, M. A. Reveley, and L. Henderson, “Saccadic abnormalities in psychotic patients.I. Neuroleptic-free psychotic patients,” Psychological Medicine, vol. 25, no. 3, pp. 461–471, 1995. [CrossRef]
23. J. A. Sweeney, M. H. Strojwas, J. J. Mann, and M. E. Thase, “Prefrontal and cerebellar abnormalities in major depression: evidence from oculomotor studies,” Biological Psychiatry, vol. 43, no. 8, pp. 584–594, 1998. [CrossRef]
24. D. J. Kupfer and F. G. Foster, “Interval Between Onset of Sleep and Rapid-eye-movement Sleep as an Indicator of Depression,” The Lancet, vol. 300, no. 7779, pp. 684–686, Sept. 1972. [CrossRef]
25. J. H. Mackintosh, R. Kumar, and T. Kitamura, “Blink rate in psychiatric illness,” The British Journal of Psychiatry, vol. 143, no. 1, pp. 55–57, 1983. [CrossRef]
26. L. R. Young and D. Sheena, “Survey of eye movement recording methods, ” Behavior Research Methods & Instrumentation, vol. 7, no. 5, pp. 397–429, Sept. 1975. [CrossRef]
27. D. Wineld and D. J. Parkhurst, “Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Workshops. vol. 3, pp. 79–79, IEEE.
28. Ioana Bacivarov, Mircea Ionita, and Peter Corcoran, “Statistical models of appearance for eye tracking and eye-blink detection and measurement,” IEEE Transactions on Consumer Electronics, vol. 54, no. 3, pp. 1312–1320, 2008. [CrossRef]
29. A. Beck, R. Steer, R. Ball, and W. Ranieri. Comparison of beck depression inventories-ia and-ii in psychiatric outpatients. Journal of Personality Assessment, 67(3):588–97, December 1996. [CrossRef]
30. Dadiz, B. G., & Marcos, N. (2018, July). Analysis of Depression Based on Facial Cues on A Captured Motion Picture. In *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)* (pp. 49-54). IEEE. [CrossRef]

AUTHORS PROFILE

Bryan Dadiz This research introduction highlights the works of Bryan Dadiz, covering various topics in computer science and technology. Dadiz has contributed to studies on emotion recognition through facial expression analysis, depression detection using video analysis, and data mining for fire incident reports. He has also developed applications for motion gesture detection and attendance tracking. Dadiz has worked on projects in artificial intelligence, machine learning, and geoinformatics, among others. His research output includes publications in international conferences and journals, showcasing his expertise in the field of computer science.

Alfio I. Regla is a information technology researcher who has published various studies that focus on developing computing solutions for different industries. His research works encompass different areas such as healthcare, education, business, and sociology. He has worked on creating solutions for hospital patient admissions, courier management, and monitoring systems for HIV patients. Regla's research also includes performance analysis of cryptographic algorithms used in the Internet of Things and research network analysis in higher education institutions. Overall, his research aims to provide practical and innovative solutions that could improve different aspects of society using computing technologies.