IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TO PREVENT THE PRISSIONERS TO COMMIT SUICIDE BY FACIAL EXPRESSION WITH THE HELP OF HUMANOID ROBORT AND DIFFERENT TOOLS &METHEDOLOGY

Alka mishara 2Vandana pathak

1M Tech Scholar, Industrial Automation and Robotics,

2Head Of Department, Industrial Automation and Robotics

1Ambalika Institute of Management and Technology, Dr A.P.J Abdul Kalam Technical University,Lucknow,

2Ambalika Institute of Management and Technology, Dr A.P.J Abdul Kalam Technical University,Lucknow

Abstract—

Recognizing human emotion is a complex task, which sometimes is necessary to help solve many related issues by robots or machines. Especially for the differently-abled persons and se- nior citizens, if their state of emotion can be recognized, the machine/robots can interact with them and help them effectively solve some problems associated with their daily lives. Also, in the arena of professional conversations and meetings, which have gone online during the COVID-19 pandemic, if we can recognize the emotion of the meeting participants through fa- cial expressions, gestures, text, and speech, the purpose of such meetings can be better served. Therefore, the present research focuses on solving such issues related to recognizing and pre- dicting human emotions by optimizing multiple modes. We develop several models and vali- date them through rigorous experiments using a real-time testbed of a humanoid robot, NAO, available at the Center of Intelligent Robotics, IIIT-Allahabad. Every mode of communica- tion is first dealt with individually, and then fusion of all the best models is done and finally improved to get the best accuracy for prediction. The modes of communication used are fa- cial expressions, the context of the image (everything other than the face), audio, and text. In another attempt to learn the dimensions of emotion, we have classified emotion dimensions based on arousal, dominance, and liking on physiological signals of the human brain.First, fa- cial emotion recognition is built and optimized, inspired by the Inception model, and iterative improvements are made by use of the Inception mechanism [1], Separable Convolution [2], Global average pooling to reduce the trainable parameters to the extent of 67%. Performance has been considerably improved with the existing state-of-the-art models, and humanoid robot, NAO is used to verify the result. So real-time implementation impact can be considerable in case of other modes of communication.While procuring physiological signals for predicting dimensions of human emotion has its societal impact, privacy hindrance is a problem and hence might be unwelcoming by the subject, for which necessary regulations and laws need to be defined which would be acceptable globally. However, this is beyond the purview of the present research which is primarily concerned with the related technology development issues only.

Index Terms: Face Recognition, Emotion Detection,Artificial Intelligence, Internet of Things,

Introduction :-

As humans, we perceive others' emotions through various cues, including visual, conversational, and contextual information. We assess facial expressions, body language, behavior, and social interactions to understand a person's emotional state. With the growing prevalence of automation and artificial intelligence, it is crucial for robots to be able to recognize and respond to human emotions effectively. This would make the interaction more user-friendly and personalized, as the robots could tailor their responses based on the context and the person's emotional state, leading to more empathetic and efficient communication. However, relying solely on facial expressions to determine a person's emotional state can be ambiguous and may lead to inaccurate predictions at times. Misinterpreting a person's emotions could result in the robot exhibiting an unwelcoming or inappropriate response, which could negatively impact the human-robot interaction In an era of autonomous and intelligent working robots entering in per- son's daily life, there is a need to bridge the gap between Artificial Intelligence algorithms and their real-time implementation ability. Robots and ma- chine intelligence are not evolving to take away employment. Instead, machine intelligence is evolving to assist human life so that manpower can be better utilized more progressively. Robots can work intelligently to assist humans in their daily activities, and machine intelligence can achieve this by training the robot to make unforeseen decisions. Also, robots can help humans replace them to perform life-threatening tasks like bomb diffusion and hard manual labor, which in earlier days cost us human lives. On the one hand, where robotic manipulators can accomplish the task of grasping and help to lift heavy objects and pick and place tasks, mobile robots can help with mobility and navigation; on the other side, humanoid robots can walk, talk and behave like human to accomplish the task which we do/ perform on a daily routine. Collaborative behavior of robots is required to adapt robots to our day-to-day activities in almost all domains.

Humanoid robots need the ability to make quick, adaptive decisions and maintain a homeostatic balance that meets human needs and makes people feel comfortable in their presence. Homeostatic balance means the robots behave in a way that aligns with human psychology. When interacting with people's daily routines, robots must have an etiquette for natural, unobtrusive behavior so that humans feel at ease in a collaborative environment. This etiquette should involve robots performing simple, supportive tasks like passing tools, providing basic customer service, and engaging in expressive communication - though the latter can be very complex to express or predict. Overall, the goal is for robots to work seamlessly alongside humans in a collaborative workspace

RESERCH AND METHEDOLOGY:

The identification of human facial emotions is a crucial aim in the contemporary technological sphere. Robotic applications are now prevalent in nearly all sectors, emphasizing the significance of emotion recognition for successful human-robot interaction. This project is focused on developing and executing a new, automated system for emotion detection and facial recognition based on Artificial Intelligence (AI) and the Internet of Things (IoT). Key terms include Face Recognition, Emotion Detection, Artificial Intelligence, and Internet of Things.

Our objective is to develop a system that utilizes machine learning and the Internet of Things (IoT) to intelligently and efficiently identify faces. Conventional face recognition systems necessitate manual input, which can be time-consuming, prone to errors, and exhibit low accuracy. By leveraging the increasing adoption of IoT devices and machine learning techniques, we can design a system that accurately recognizes individuals without requiring user input, by learning from facial feature patterns. This system can find applications in attendance tracking, security systems, and personalized marketing. Nevertheless, the challenge lies in creating a system that is both trustworthy and secure, while simultaneously safeguarding the privacy of users The aim of implementing smart facial recognition technology using IoT and machine learning is to enhance the security and efficiency of different businesses and public areas. By utilizing cameras and sensors, this technology captures images and videos of individuals, which are then analyzed by machine learning algorithms to identify faces and compare them against a database of known individuals.

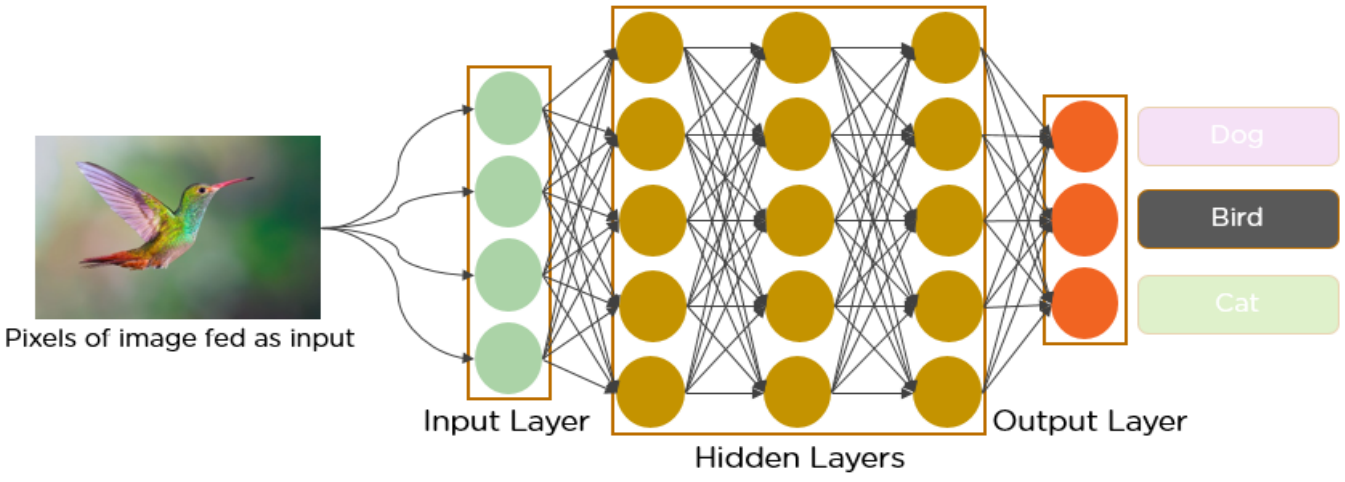
Integrating smart facial recognition can help organizations improve security measures by monitoring employee attendance, detecting unauthorized access, and identifying potential risks promptly. This technology also enables quick and easy access control to restricted areas such as banks, airports, and government facilities.

Moreover, IoT-enabled smart facial recognition can enhance the consumer experience in various sectors like hospitality, retail, and healthcare. By analyzing clients' faces and preferences, services and recommendations can be personalized accordingly. The use of IoT and machine learning in developing smart facial recognition aims to boost security, efficiency, and customer satisfaction across different commercial sectors and public spaces.

CONVOLUTION NEURAL NETWORK:-

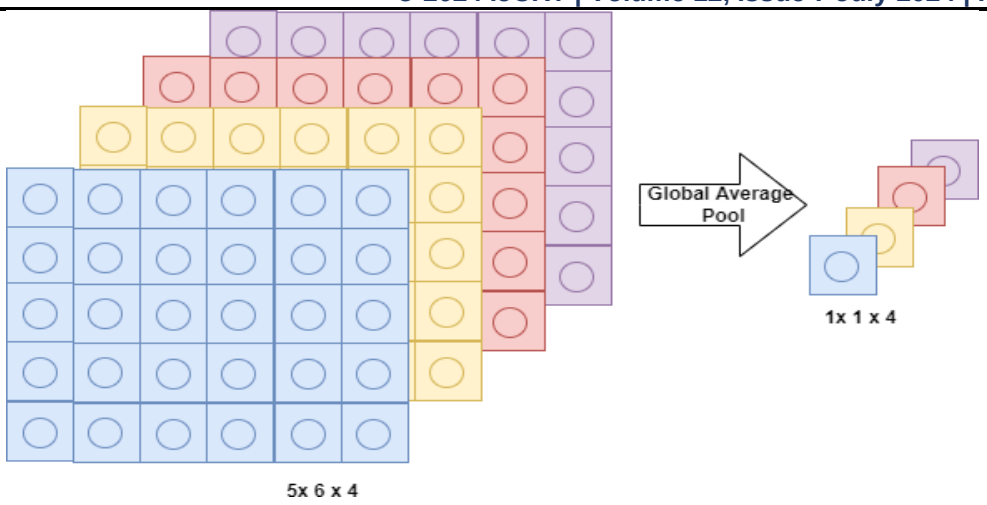
Convolution Neural Network (CNN) is a variant of Artificial Neural Network where all the neural connections are replaced by convolution operation via fil- ters. CNN further utilizes the benefits of parameter sharing and translation invariant, resulting in reduced computation (as compared to Artificial Neural Network) and improved performance. Convolution operation is the most appropriate way to figure out the edges and obtain a feature map to best classify the image.

To process the image dataset, a convolution neural network was utilized to extract features in the form of edges for classification tasks without relying on human-predicted features. Convolutions facilitate parameter sharing and translation invariance, enabling the detection of features with reduced computation compared to fully connected networks. The convolution neural network is considered one of the most effective networks for processing static images due to its accuracy, computational efficiency, ability to handle images of varying sizes, adaptability to dataset changes without network modification, and batch data processing capability. Unlike other networks such as multi-layer perception and recurrent neural network, CNN requires minimal preprocessing for managing diverse datasets in terms of size and quality. It comprises Convolution layer, Pooling layer, ReLU layer, Fully Connected Layer, and Loss layer. The convolution layer conducts the primary convolution operation to extract edges, with trainable filters that can be frozen if necessary. Given the high dimensionality of images, connecting them through a neural network is impractical, which is addressed by CNN through parameter sharing and local connectivity for computational benefits.



GLOBAL AVERAGE POOLING:-

GAP aggregates the average value of a feature map obtained after the convolution operations, significantly reducing the parameters compared to fully connected layers. This helps prevent over fitting, which could occur due to excessive parameter learning. The original concept of GAP, proposed in [83], suggests replacing fully connected layers entirely. Once the convolution layers obtain a feature map, GAP takes the average of each feature map and directly feeds it to the softmax layer to determine the dominant class label, thus predicting that class. This approach directly corresponds to the class label from the feature map. Additionally, Global Average Pooling is more invariant to feature translation, extracting the dominant feature to determine the predicted class. Overall, it reduces computational complexity by a large extent, up to 80% depending on the network architecture



Spectrogram Image based Emotion Recognition in Audio-

Recognizing human emotions is a challenging task that has been the subject of research for many years. The issue remains relevant due to its importance in various fields related to human-computer or robot interaction. According to studies, individuals can gauge the emotional states of others by observing a variety of parameters, with 70% of these parameters being non-verbal. Emotions are conveyed through speech, posture, gestures, context, facial expressions, and even the history of a conversation or situation. These individual components can be effectively addressed using learning-based methods. Predicting emotions in multi-party audio conversations adds complexity to the problem, as it requires understanding speech intent, cultural nuances, accents, gender differences, and other factors. Researchers have endeavored to categorize human audio into specific classes using the Support Vector Machine model, Long Short Term Memory (LSTM), and bi-LSTM on audio inputs. Our proposed approach involves using an image-based emotional classification method for audio conversations. By converting the spectrogram of an audio signal into an image and inputting it into a Convolutional Neural Network model, we can extract patterns for classification. This approach has shown promising results, achieving an accuracy of approximately 86% on the test dataset, which represents a significant improvement over existing models.

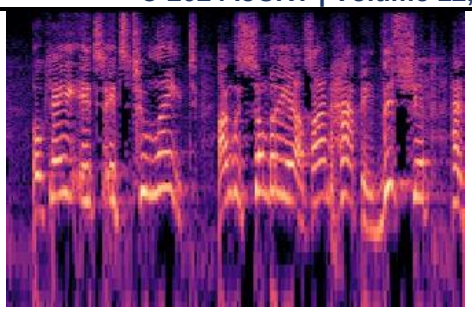
Dataset Description:-

The dataset utilized in this study is the Multimodal Emotion Lines Dataset (MELD), which includes text data along with audio and visual components.

This section focuses on the Audio Dataset, where 13,000 audio recordings are categorized into seven emotion classes - joy, anger, surprise, neutral, sad, disgust, and fear. The MELD dataset comprises video recordings and dialogue utterances in the text format. We have extracted the audio from the video recordings in CSV format and transformed it into spectrogram images for our experiments. Ultimately, we obtained 9989, 1109, and 2610 images for the training, development, and testing data of the audio files.

Data Preprocessing

1. Given that a Convolution Neural Network model was utilized for classification, the audio files were required to be organized into seven folders representing their respective classes. This task was accomplished by leveraging metadata in a CSV file in conjunction with a Python script. Python served as the programming language for conducting the experiments. The audio recordings underwent conversion into Spectrograms using the Python Librosa Library. The time v/s frequency plot axis was disregarded, and the spectrogram images were cropped as per the specified criteria. The cropped Spectrogram image of an audio recording categorized under the 'anger' class is displayed. These cropped spectrogram images were inputted into the CNN model for the training process.



Spectrogram of Emotion - Anger

7.1.1 Flow of Proposed approach

Deep learning has experienced significant achievements in recent years, primarily due to the abundance of available datasets for training purposes. The task of emotion recognition has always been a challenging one in the realm of Machine Learning. Efficient algorithms for emotion recognition should possess the capability to accurately classify emotions to the highest degree possible.

We have introduced a novel image-based method for predicting emotional states from audio conversations in our current study. By correlating frequency modulations with emotions, we were able to identify various tone variations in emotions through spectrogram representation. This, in turn, facilitated the accurate classification of emotions into their respective categories.

Our experimentation involved utilizing training data from the MELD Dataset. Additionally, we conducted testing on the test dataset to evaluate the performance of our model further. Convolutional Neural Networks yielded superior results compared to alternative methods. The experimental findings demonstrate the model's commendable recognition accuracy.

In our future endeavors, we aim to incorporate other modalities such as text and video into the emotion recognition task, thereby elevating its complexity. Given that emotions encompass multiple sources including text, audio, video, facial expressions, and body postures, integrating all these elements into a unified framework can significantly impact real-time human-robot interaction or computer interaction.

RESULT ANALYSIS

suitable for real-world applications. The model incurs a loss of approximately 0.712, as depicted in . Validation accuracy, as shown in Fig. 4.3, stabilizes at 74%, showcasing an improvement over previous state-of-the-art research by reducing parameters by 50%. This enhancement in accuracy is achieved through adjustments in filter size, network convolution layers, and down-sampling image size in later network stages to extract optimal features across deep layers.Utilizing diverse datasets enhances the network's robustness by exposing it to various variations such as non-face images, different lighting conditions, age differences, occlusions, makeup, diverse backgrounds, and a range of facial variations. Upon testing the network across all mentioned datasets, the confusion matrix for the JAFFE dataset is illustrated .The analysis reveals that most false predictions occur in disgust emotions, often misclassified as sad or neutral, mirroring human predictions. The next highest false predictions are in the neutral faces of the JAFFE dataset, where neutral expressions are occasionally misinterpreted as sad, surprised, or fearful, attributed to the network's interaction with Japanese Itdisplays a selection of images from the FER2013 dataset, showcasing variations in age groups, lighting effects, and The network has demonstrated robustness in emotion recognition, operating at real-time speed and being facial angles. This dataset encompasses a wide range of face and non-face images to facilitate the network's generalization. Additionally, Fig. 4.6 presents results from the JAFFE dataset, highlighting true positives and false positives in predicted expressions. Finally,illustrates the impact of images from a custom dataset on predicted labels, demonstrating the network's ability to detect faces.

CONCCLUSION AND FUTURE WORK:-

Our research is a summary of the increasing need for robots, especially humanoid robots, to understand human emotions in order to enhance human-robot interaction. We have developed a real-time robust emotion classifier that reduces computational cost and achieves near-human accuracy. This model, built with 146,000 parameters, took eight hours to train and achieved an accuracy of 65% on training and 74% on validation. The robustness of the model was verified over eight datasets for emotions. By incorporating such an emotion classifier, personal robots can become more familiar and customized to the user, leading to a more realistic connection between the person and the robot. Our proposed method also reduces parameter requirement, memory requirement, and computation cost. In the future, we plan to integrate this model with the NAO robot, a humanoid robot, to make them more social and realistic in their interactions with humans. Additionally, we aim to integrate more emotions that humans can understand, distinguishing mixed feelings to make conversations more realistic and customized. The robot must interpret emotions within text and understand the context to grasp a person's sentiment accurately. Recognizing facial expressions is also crucial. Therefore, a model is suggested to forecast facial expressions in real-time to enhance humanoid robots' understanding of individuals. This research will enhance the connection between robots and humans, allowing humans to feel more at ease in the presence of robots. Consequently, humans can communicate more openly with robots and establish a more personalized relationship. As mentioned in [127], the authors have explored various methods of emotion prediction, which can be beneficial in fields such as image processing, cybersecurity, robotics, psychological research, and virtual reality applications, including robots' social interaction with humans.

The proposed model has been tested with a humanoid robot, resulting in improved responses. The suggested network has reduced the parameter requirement by 94%, reducing complexity. An overall accuracy improvement of up to 6% has been achieved when implemented on humanoid robots, compared to latency and response time. There has been enhancement in real-time systems, specifically humanoid robot, NAO.
By integrating humanoid robots such as NAO, the suggested network system is capable of assessing an individual's emotional state and responding appropriately. Additionally, we can evaluate the system's effectiveness by comparing it with other social robots. Enhancing the robot's ability to communicate through multiple modalities could also make it more sociable and easier to understand. The Affectnet dataset includes emotion categories like excitement, joy, contentment, and no emotion, among others, which could be further integrated into the system. The emotional response to a statement can vary across different times and contexts, but with facial emotion recognition, these variations can be accurately predicted. Furthermore, researchers [128] have identified a range of complex emotions such as "angrily disgusted," "sadly angry," etc., which could be incorporated to improve the system's ability to predict and facilitate more effective collaboration.

In future research, the system's performance in humanoid robots could be enhanced by incorporating data from both NAO's cameras to boost prediction accuracy. [103] have demonstrated the ability to predict finger-pointing direction using a single RGB camera. Since the robot is already equipped with an emotion classifier, integrating gestures and speech could make it more natural and effective. Consequently, personal robots could be utilized in various roles such as psychological counseling, child care, and elderly care.

However, the work is not without its limitations. The system could benefit from the addition of more complex emotions to improve prediction accuracy. Moreover, exploring alternative communication methods could further enhance the system's performance. Throughout the current investigation, we have endeavored to enhance the efficiency and decrease the computational complexity of emotion classification problems across various modes of communication. Our efforts have focused on reducing the parameters needed for model training, particularly in the realm of facial emotion recognition, in order to enhance accuracy. We have integrated a range of methods and mode mechanisms into our experiments, such as Separable convolution, Inception mechanism, and batch normalization.

We have carefully balanced the reduction in trainable parameters with accuracy, ensuring that we do not sacrifice accuracy in order to reduce computational complexity. Our facial emotion recognition model, along with several baseline models, has been implemented on the humanoid robot, NAO, in a real-time testbed to conduct a comparative study on performance improvement in real-time. Additionally, we have incorporated context information alongside facial data to enhance emotion prediction performance, and have conducted an analysis to assess the relevance of context information in emotion recognition of humans. Furthermore, we have utilized audio and text data to predict human emotions. Finally, we have proposed a fusion model to predict human emotion, incorporating face, context, audio, and text modalities. Our experiments have focused on creating a fusion function for different modalities, ensuring that the generality of each modality is not impacted by features of other modalities, and that the prediction is made using the remaining modalities even if one is missing. We have also considered the correlation between the modalities and improved the prediction performance. As the docking layer in EmbraceNet takes features in a multinomial distribution fashion, it implements dropout operation, thereby preventing overfitting in the model. In future work, we aim to implement the model in a real-time test bed to compare and work on the performance of a real robot. Additionally, the model will be further developed to...

REFERENCES

1. [TP91] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71-86, 1991.

2. [ZKC+98] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets and J. Weng.Discriminant analysis of principal components for face recognition, pages 73-85. Springer Verlag Berlin, 1998.

3. [GHW12] M. Günther, D. Haufe and R.P. Würtz. Face recognition with disparity corrected Gabor phase differences. In Artificial neural networks and machine learning, volume 7552 of Lecture Notes in Computer Science, pages 411-418. 9/2012.

4. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah and S. U. Ali, "Facial Recognition usingConvolutional Neural Networks and Implementation on Smart Glasses," 2019 International Conference on Information Science and Communication Technology (ICISCT), 2019, pp. 1-6, doi: 10.1109/CISCT.2019.8777442.

5. Mehedi Masud, Ghulam Muhammad, Hesham Alhumyani, Sultan S Alshamrani, Omar Cheikhrouhou, Saleh Ibrahim, M. Shamim Hossain,

6. Deep learning-based intelligent face recognition in IoT-cloud environment,

7. Computer Communications, Volume 152, 2020, Pages 215-222, ISSN 0140- 3664.

8. Bhatti, K., Mughal, L., Khuhawar, F., &Memon, S. (2018). Smart attendance management system using face recognition. EAI Endorsed Transactions on Creative Technologies, 5(17).

9. Kumar, P. M., Gandhi, U., Varatharajan, R., Manogaran, G., & Vadivel, T. (2019). Intelligent face recognition and navigation system using neural learning for smart security in the Internet of Things. Cluster Computing, 22(4), 7733-7744.

10. Agarwal, L., Mukim, M., Sharma, H., Bhandari, A., & Mishra, A. (2021, March). Face recognition based smart and robust attendance monitoring using deep CNN. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 699-704). IEEE.

11. Kumar, T. A., Rajmohan, R., Pavithra, M., Ajagbe, S. A., Hodhod, R., & Gaber, T. (2022). Automatic face mask detection system in public transportation in smart cities using IoT and deep learning. Electronics, 11(6), 904.

12. Atik, M. E., & Duran, Z. (2020, October). Deep learning-based 3d face recognition using derived features from point cloud. In The Proceedings of the Third International Conference on Smart City Applications (pp. 797-808). Springer, Cham.

13. BoserB ,Guyon I.G,Vapnik V., "A Training Algorithm for Optimal Margin Classifiers", Proc. Fifth Ann. Workshop Computational Learning Theory,pp. 144-152, 1992.

14. Mitchell, T. (1997). Machine Learning, McGraw Hill. ISBN 0-07-042807-7., McGraw-Hill, Inc. New York, NY, USA. Published on March 1, 1997

15. Alex C, Boston A. (2016).Artificial Intelligence, Deep Learning, and Neural Networks, Explained (16:n37)

16 Varun G., Lily P., Mark C., "Development and validation of a deep learning Algorithm for Detection of Diabetic Retinopathy", December 2016.

17 Tiago T.G. "Machine Learning on the Diabetic Retinopathy Debrecen Dataset", knowledge-Based System60, 20-27. Published on June 25, 2016.

18. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556–64

19. Boser B. E, Guyon I. M. and Vapnik V. N. (1992). "A training algorithm for optimal margin classiers".Proceedings of the 5th Annual Workshop on Computational Learning Theory COLT'92, 152 Pittsburgh, PA, USA. ACM Press, July 1992.