International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

Obstructive Sleep Apnea Prediction Using Deep Learning Techniques

A. Pavani[1], Kalyanapu Srinivas[2]

[1]M.Tech Student, [2]Associate Professor, Dept. of Computer Science & Engineering, Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India.

ABSTRACT - Obstructive sleep apnea (OSA) is a medical condition in which the airway becomes obstructed regularly, and resulting in sleep disruption. Sleep apnea is a breathing condition in which a person stops breathing repeatedly while sleeping. OSA is usually diagnosed by an expensive procedure that requires the patient to remain in the hospital overnight. However, this method is expensive, inconvenient, and time consuming. Snoring, poor night sleeps due to choking or gasping, and waking up unrefreshed are all common symptoms of OSA. In this paper, we proposed the two deep pre-trained convolutional neural network models ResNet50 and InceptionV3 for predicting obstructive sleep apnea. We consider the 2-D facial images of different subjects of various conditions such as sleepy, normal, happy and sad images. Even with the small amount of data using transfer learning, InceptionV3 had a good average accuracy of around 91% for OSA prediction and ResNet50 had a good average accuracy of around 82%.

Key Words: Obstructive Sleep Apnea, Deep Learning, Transfer Learning, ResNet50 model, InceptionV3 model.

I. INTRODUCTION

Obstructive sleep apnea (OSA) is major type of sleep-related breathing disorders. Poor sleep has a major impact on social and personal activities. Due to obstructions in the airway, it affects a person's ability to sleep well, causing loud snoring, choking, and gasping for breath while sleeping, as well as daytime sleepiness and headache. If not addressed, this could lead to health complications including hypertension, heart attacks, diabetes, depression, and early death [1].

Sleep apnea is classified into three types: Obstructive sleep apnea (OSA), which is the most popular form of the disease, central sleep apnea (CSA), and mixed sleep apnea (MSA). The first type is caused by the tongue or upper airway muscles relaxing while sleeping, obstructing the airway and making breathing difficult. The second type happens when the brain fails to send proper signals to the muscles to breathe. The third type is comprised of symptoms from the first and second combined. OSA is more common than Central Sleep Apnea.

Breathing disorders occur as irregular events, the most common of which are apnea and hypopnea. A complete or nearly complete absence of airflow for yet exceeding ten seconds is defined as an apnea event, while a hypopnea event is defined as a reduction of airflow for yet exceeding ten seconds [2]. Our upper airway normally remains open during sleep because the muscles lining the upper throat are relaxed. However, in OSA, a person may have a recurring obstruction in the upper airway for yet exceeding ten seconds for various reasons, which causes the lungs out of oxygen and the person to wake up, which would restore the airway. If more than 15 apneas occur then the analysis of OSA is made.

Obstructive sleep apnea is diagnosed using the patient's medical history, physical examination, polysomnography (PSG) test, and scanning. The gold standard to diagnoses is PSG test, in which patients are examined overnight with various sensors connected to their bodies. Signals from different sources, such as oxygen level, heart rate, and body movements, are measured by these sensors. PSG recordings are analysed by sleep specialists, who manually recognize apnea and hypopnea events. The number of these irregular events occurs during the night determines the severity of sleep apnea. This procedure has many limitations such as (i) requires a special lab for testing, (ii) more connected wires, (iii) the use of a nasal air flow tube (iv) During sleep, patients do not move, (v) creates anxiety in the patient, which can have an effect on the actual OSA result, (vi) requires a sleep expert to read the data and connection of sensors, and (vii) PSG is expensive and time consuming [3].

The advanced applications of Deep Learning are discussed in this paper. Deep convolution neural network has made a significant contribution in a number of areas, including image classification and recognition. Then we discuss the transfer learning of a pre-trained convolution neural network models like ResNet50 and InceptionV3.

 DOI : 10.35291/2454-9150.2021.0288

The paper is presented as follows: In Section 2, we glance at a variety of OSA detection methods. Dataset in Section 3.1, Deep learning, convolutional neural network, transfer learning pre-trained models is described in Section 3.2, 3.3 and 3.4, respectively. Proposed Methodology describes the Section 4. Experiment, Results and Discussion are examined in Section 5 & 6. Finally concludes the paper in Section 7 and Future Work in Section 8.

II. RELATED WORK

Imaging methods used to make diagnosis simpler and faster. The authors used sophisticated volumetric analysis on MRI [4]. They found physiological variations in tongue volume and with soft tissues using statistical tests. Lowe et al. [5] the impact between craniofacial structure measured by lateral cephalometry and size of the upper airway, tongue, and soft palate determined by computed tomography scans. To advance our understanding of the pathogenesis of OSA and establish a prediction equation for apnea severity, researchers used cephalometry and three-dimensional upper airway CT.

R. W. W. Lee et al. [6] study the impact between upper airway structures and surface facial measurements with OSA using MRI. Dimensions of facial surface (such as facial widths and heights, and nose width) have significant positive associations with structure of upper airway. The major face variations were discovered in the lower part of face and the upper part of the neck [7]. These findings prompted researchers to develop a facial image-based OSA diagnosis. However, because of the radiation exposure, these treatments are generally expensive, time-consuming, and invasive for patients.

The rationale of using speech and image analysis in OSA assessment can be found in works like Lee et al. [8], [9], have used features of craniofacial surface structures (such as the head, eyes, nose, face, and neck) from digital photographs to predict OSA. While excessive weight and an excess of regional adipose tissue are major risk factors for OSA, craniofacial abnormalities and an altered upper airway structure are also significant interacting factors in OSA pathogenesis. They analyzed frontal and profile photographs of 114 subjects. They achieved a 76.1% accuracy using photographic measurements and logistic regression, and the Receiver Operating Characteristic (ROC) curve of 0.82.

Also [10], [11] have explored the prediction capability of OSA by analyzing facial profile and frontal images. Supervised automatic image processing is used instead of precise manual identification for the purpose of landmark identification. AHI prediction was also evaluated using clinical variables such as height, weight, age, BMI (Body Mass Index), and neck circumference, in addition to facial and voice features. These findings demonstrate that features capturing the composite elements of craniofacial structures and regional adiposity can better predict OSA than demographic data gathered from clinical observations. However, since digital photographs are 2D, non-linear measurements and shape of craniofacial anatomy are obtained. All of these methods rely on manual landmark detection, which takes time and is also dependent on the experience of the person who will mark the key point on the facial data.

The probability of automated detection of OSA using front profile face images was investigated by AT Balaei et al. [12]. First, the face images with their manually annotated landmarks were used to find out the facial landmarks in the new images automatically. To predict OSA, these landmarks were used in a classifier. Using a logistic classifier, an accuracy of 70% was achieved in this step. The same training images (after pre-processing) were fed directly into a neural network classifier in a second stage, yielding an accuracy of 62% in detecting OSA.

Rim Haidar et al. [13] using CNN and markov chain predicting sleep apnea event, A large dataset with 48,000 examples from 1,507 subjects, the proposed system for automatically learning the necessary features and predicting sleep apnea event's. The results show that, an accuracy of 80.78% and an F1 score of 80.63%. The Markov chain rules were examined, and provide an overview of the transitions between apnea and normal events.

The authors proposed a new approach to detection of apnea events using CNN and the raw signal from nasal airflow [14]. In order to classify an apnea event, they used two methods: Convolution neural network (CNN) and support vector machine (SVM). The back propagation algorithm and the Adam optimizer were used to train the neural network. In this analysis, data from 100 patients was used. The ability to learn information features from high-dimensional data no need of signal processing and feature engineering is the key advantages of using CNN and SVM. The results of the experiment revealed that CNN outperformed the SVM in their system, with an accuracy and F1-score of about 75%. However, the proposed device has drawbacks, including the possibility of the airflow sensor detachment from a set position due to the patient moving during sleep, and the patient breathing through the mouth rather than the nose.

Syed MS Islam et al. [15] investigated a method for the detection of OSA by using a deep learning technique through depth maps of human facial scans. Artec Eva used Artec Studio to create facial depth maps from 3D scans. MeshLab software was used to generate a 2D image after converting the facial depth map to eliminate unnecessary variations. A small sample of 69 volunteers was collected from Genesis Sleep Care. As a result, three transfer learning techniques were used: face recognition by the

 DOI : 10.35291/2454-9150.2021.0288

ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

Visual Geometry Group (VGG), pose-aware models-AlexNet (PAMs-AlexNet), and PAMs-VGG19. The proposed framework for OSA prediction is developed in MATLAB and is based on end-to-end deep learning. The best performance of VGG Face was found 68.75% (validation) accuracy and 67.42% (test). The study's main drawback was the limited number of participants, which made it difficult to validate and evaluate the deep learning technique. This system also has expensive image processing components, and the detection accuracy is relatively low.

Our goal is to detect OSA without human intervention by using 2D information from facial features and developing a framework such as MATLAB Deep Learning Toolbox. For that, we have proposed an Obstructive sleep apnea prediction using a deep convolution neural network based on pre-trained transfer models. To achieve higher accuracy, we used ResNet50 and InceptionV3 pre-trained models.

III. MATERIALS AND METHODS

3.1 Dataset

Two-dimensional images are collected from the AR Face Database [16], which includes 60 images of 15 different subjects in four different conditions: Sleepy, Normal, Happy, and Sad.

3.2 Deep Learning

Deep learning is a sub-field of machine learning that is influenced by the brain's structure. In the area of medical image processing, as in many other areas, deep learning methods have continued to produce promising results in recent years. Deep learning models are used to analyse image and signal data collected using medical imaging techniques such as magnetic resonance imaging (MRI) [19], computed tomography (CT), and X-ray. The result of these studies, disorders such as diabetes, brain tumours, skin cancer, and breast cancer can be detected and diagnosed more easily.

Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained using a wide collection of labelled data and multilayer neural network architectures. Machine learning is a form of deep learning. Deep learning models, often exceeding human-level efficiency, may achieve state-of-the-art precision. A wide collection of labelled data and neural network architectures that include several layers, models are trained. The number of hidden layers in a neural network is generally referred to as "deep". Without involving an objective researcher, the deep learning network's hidden layers implicitly perform all of these tasks.

3.3 Convolutional Neural Network

A convolutional neural network (CNN) introduced by Fukushima in 1998, CNN are seen in an activity recognition, sentence classification, and text recognition, face recognition, object detection and localization, image characterization etc. A CNN [20] is a Deep Learning algorithm that accepts an image as input, assign importance to different aspects of the image, and differentiate between them. The main building blocks of CNN are convolution layer and pooling layer, followed by fully connected layers and an output layer as seen in Fig 1.

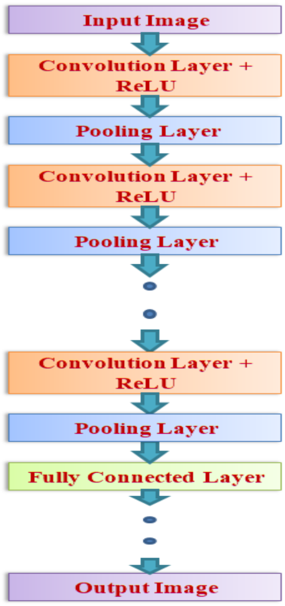

Fig 1: Building Blocks of CNN

Input layer: An Image is given as input; this layer is responsible for reading the pre-processed image dataset. Cropping and resizing the images are part of the pre-processing phase.

Convolutional Layer: Convolution layer is responsible for determining the features of the pattern. In this layer, the input image is passed through a filter. The values resulting from filtering consist of the feature map. The kernel (or weight) is a 3x3 or 5x5 shaped matrix is used to extract low-level features from the original image matrix. The role of the filters are extracting some features from the input images and grouping them in feature maps. Stride parameter is the number of steps tuned for shifting over input matrix. During the convolution process, the weights are learned so that the loss function is minimised.

ReLU layer: A ReLU implements the function **y = max(x, 0).** The use of ReLU layer is to replace the negative pixel values by zero. This is to produce a non-linearity map of features on the CNN network.

Pooling layer: Pooling means down sampling of an image. It sub-samples a small region of the convolutional output as input to produce a single output. The largest of a region's pixel values is used in max pooling. The other intermediate layer used is the dropout layer. The main

 DOI : 10.35291/2454-9150.2021.0288

purpose of this layer is to prevent network overfitting and divergence.

Fully Connected Layer: Fully connected is the last and most important layer of CNN. This layer functions as a multi-layer perceptron. To generate output, this layer uses input from all the neurons in the previous layer and performs an arithmetic sum of the previous layer's features for each neuron in the current layer.

Output Layer: The final layer is Output layer is also called as Softmax layer. The Softmax function is used to generate probabilities for each target variable class.

3.4. Transfer Learning Pre-Trained Models

One of the most difficult challenges researchers' faces when analysing medical data is the limited number of datasets available. Deep learning models frequently require a huge amount of data. Labelling this data by experts is complex and expensive. The most significant benefit of using the transfer learning approach is that it allows for data training with less datasets and lower computation costs. With the transfer learning method, which is widely used in the field of deep learning, the information gained by the pre-trained model on a large dataset is transferred to the model to be trained. We will identify some of the current state-of-the-art deep transfer learning models in this section.

A. ResNet50

ResNet50 is a pre-trained model that won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) competition in 2015. A subset of the ImageNet database was used to train it. The model was trained on over a million images, has 177 layers (equivalent to a 50-layer residual network), and can classify images into 1000 different object categories. ResNet stands for Residual Network, which is a network that supports Residual Learning. The number 50 denotes the number of layers. Resnet50 refers to a 50-layer residual network. An overview of ResNet50 architecture as shown in Fig. 2

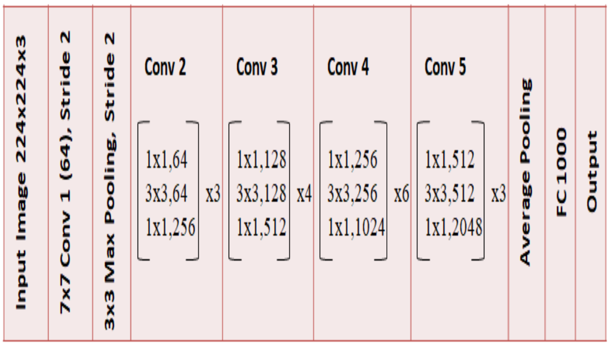

Fig 2: ResNet50 Architecture

As the network depth increases, accuracy grows saturated and then rapidly reduces. By using skip connections, the Microsoft Research team concentrated on this issue with ResNet. Batch normalisation was first introduced in the ResNet50 architecture [17]. The vanishing gradient problem affects the deeper CNN with more layers. It affects the deeper CNN with more layers. Identity mapping is performed using a pre-trained model with additional layers to solve this problem. As a solution to the degradation problem, the deep residual learning scheme is introduced.

B. InceptionV3

InceptionV3 [18] is the first runner up for image database in 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). InceptionV3 is GoogleNet's upgraded version, which reduces computational complexity. InceptionV3 aimed to improve the use of resources within the network by increasing the depth and width of the network while maintaining constant computing operations. InceptionV3 yields highly accurate results. Fig. 3 shows an overview of InceptionV3 architecture.

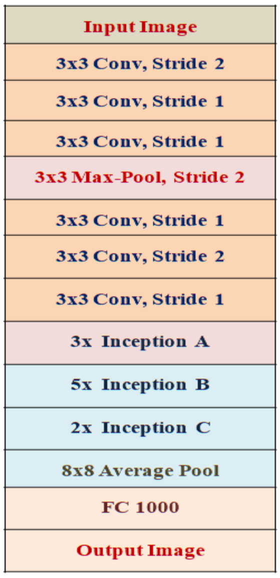

Fig 3: InceptionV3 Architecture

The term "inception modules" is to describe an optimized network structure. To minimise dimensionality to a manageable computing standard, this inception module is repeated spatially by stacking with occasional max-pooling layers. The InceptionV3 network is made up of a number of symmetrical and asymmetrical building blocks, each with multiple convolutional branches, average pooling, max pooling, dropouts, and fully-connected layers.

IV. PROPOSED METHODOLOGY

Sleep data were collected in the form of 2-D images. The input to our model is 2D images. Sleep data information to detect OSA in different subjects without human intervention. To train a CNN from scratch needs a huge amount of data, this in our case is very less. For that we

 DOI : 10.35291/2454-9150.2021.0288

trained a deep transfer learning pre-trained networks Resnet50 and InceptionV3. Transfer learning is already trained model and directly applies on test data making it suitable for small amounts of data. In our model, we used the MATLAB deep learning toolbox. Matlab deep learning toolbox is used to import the pre-trained networks ResNet50 and InceptionV3. An overview of the proposed framework is as shown in Fig 4. When there are enough examples, fine tuning will classify the network layer and improve generalisation. One type of supervised learning problem is classification. Where the machines try to predict the category or class of given input data and achieve high accuracy.

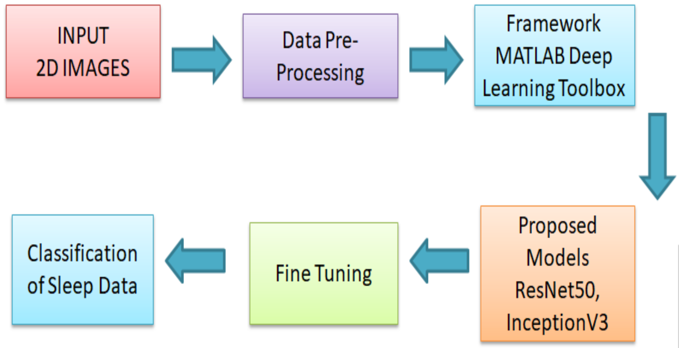

Fig 4: Proposed Framework

V. EXPERIMENT

The Experiment was carried out using the MATLAB Language, version R2018a. The System in which the experiment was carried out run on Windows 10 and has a Random Access Memory (RAM) of 512MB and Hard Disk of 40GB were used in this experiment. To run the MATLAB deep learning toolbox, we must first install two pre-trained models ResNet50 and InceptionV3, by simply creating an account in Mathworks with our email address.

A. Splitting Data

The dataset consists of 60 images of 15 different subjects categorized as Sleepy, Normal, Happy and Sad images. We are using the SplitEachLabel function to divide the data into 70% for training and the remaining 30% for testing. Training is done using the MATLAB deep learning toolbox.

B. Data Pre-processing

The input image size for ResNet50 and InceptionV3 should be of the size 224x224. Pre-processing data on image performing some geometric and color transformations like scaling, rotation, cropping and converting gray scale to RGB image. The most common method to reduce overfitting on image data is the Data augmentation.

The model must be compiled after pre-processing and before training the network. We will set the feature layer as 'Fc1000' for ResNet50 and feature layer 'Predictions' for InceptionV3. We set the batch size to 8, and we must declare a few parameters such as the optimizer, loss function, and metrics to be calculated during training. After extracting the features from the training dataset, we train a liner classifier, which is a Random forest classifier by default. Categorical-Cross Entropy is the loss function used to find the loss for this model.

C. Training the model

Training is done using the MATLAB deep learning toolbox. We are using the transfer learning pre-trained models such as ResNet50 and InceptionV3. We will train the network end-to-end classification. The predicted values are similar to the one which we're comparing, so if they're the same, we'll count as one otherwise zero and resulting overall accuracy. We will give the input as 2D images and it will classify the images and getting the accurate results.

VI. RESULTS AND DISCUSSION

A new set of parameters is used to train the network. The ResNet50 feature layer will be 'Fc1000', and the InceptionV3 feature layer will be 'predictions'. Table I show that the accuracy of two different proposed models. InceptionV3 outperforms ResNet50 in terms of accuracy. When compared to ResNet50, InceptionV3 has deeper layers. The benefit of InceptionV3 is that it has been pre-trained on a large dataset. Obstructive sleep apnea prediction using deep learning techniques, InceptionV3 performed best with an accuracy of around 91%, while ResNet50 performed best with an accuracy of around 82%.

MODELS	ACCURACY
InceptionV3	91%
ResNet50	82%

Table I: Accuracy of two different proposed models

COMPARISON:

We performed a comparison with other deep learning techniques such as VGG 16, VGG 19, and AlexNet. Table II shows a comparison of the existing and proposed methodologies. As can be seen, the InceptionV3 model outperformed all other deep learning techniques in terms of accuracy.

 DOI : 10.35291/2454-9150.2021.0288

Table II compares the existing model with the proposed model

EXISTING MODEL		PROPOSED MODEL	
NAME OF THE MODELS	ACCURACY VALUES	NAME OF THE MODELS	ACCURACY VALUES
VGG 16	81%	InceptionV3	91%
VGG 19	76%		
AlexNet	72%	ResNet50	82%

VII. CONCLUSION

In this paper, we proposed the two deep pre-trained convolutional neural network using 2D images and transfer learning models ResNet50 and InceptionV3 for predicting OSA. Even with the small amount of data, to overcome this limitation we made use of transfer learning. We analyze two pre-trained models and among them, InceptionV3 performs the best. Our method shows comparable performances to the state-of-the-art results in terms of getting prediction using end-to-end deep learning. InceptionV3 had a good average accuracy of around 91% for OSA prediction and ResNet50 had a good average accuracy of around 82%.

VIII. FUTURE SCOPE

Future work will include further evaluation of the proposed approach. The collection of more facial images improves the robustness, accuracy, and computational efficiency of using different deep learning techniques to predict obstructive sleep apnea.

REFERENCES

[1]. "Obstructive sleep apnoea (osa: sleep apnea) information", Jul 2018. [Online] Available: https://www.myvmc.com/diseases/obstructivesleep-apnoea-osa-sleep-apnea/

[2]. M. R. Mannarino, F. Di. Filippo and M. Pirro, "Obstructive sleep apnea syndrome", *European journal of internal medicine*, vol. 23, no. 7, pp. 586-593, 2012.

[3]. R. Haidar, S. McCloskey, I. Koprinska and B. Jeffries, "Convolutional neural networks on multiple respiratory channels to detect hypopnea and obstructive apnea events", *International joint conference on neural networks (IJCNN)*, IEEE, 2018, pp. 1-7.

[4]. R. J. Schwab, M. Pasirstein and R. Pierson, "Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging", *American journal of respiratory and critical care medicine*, vol. 168, no. 5, pp. 522-530, 2003.

[5]. A. A. Lowe, J. A. Fleetham, S. Adachi and C. F. Ryan, "Cephalometric and computed tomographic predictors of obstructive sleep apnea severity", *American journal of orthodontics and dento facial orthopedics*, vol. 107, no. 6, pp. 589-595, 1995.

[6]. R. W. W. Lee, K. Sutherland, A. S. L. Chanetal, "Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea", *Sleep*, vol. 33, no. 9, pp. 1249-1254, 2010.

[7]. E. K. Pae, A. A. Lowe and J. A. Fleetham, "Shape of the face and tongue in obstructive sleep apnea patients statistical analysis of coordinate data", *Clinical orthodontics and research*, vol. 2, no. 1, pp. 10-18, 1999.

[8]. R. W. W. Lee, A. S. L. Chan, R. R. Grunstein and P. A. Cistulli, "Craniofacial phenotyping in obstructive sleep apnea – a novel quantitative photographic approach", *Sleep*, vol. 32, no. 1, pp. 37-45, 2009.

[9]. R. W. W. Lee, P. Petocz, T. Prvan, A. S. L. Chan, R. R. Grunstein and P. A. Cistulli, "Prediction of obstructive sleep apnea with craniofacial photographic analysis", *Sleep*, vol. 32, no. 1, pp. 46-52, 2009.

[10]. F. Espinoza-Cuadros, R. Fernandez-Pozo, D. T. Toledano, J. D. AlcazarRamırez, E. Lopez-Gonzalo and L. A. Hernandez-Gomez, "Speech signal and facial image processing for obstructive sleep apnea assessment", *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[11]. H. Nosrati, N. Sadr and P. de Chazal, "Apnoea-hypopnoea index estimation using craniofacial photographic measurements", *Computing in cardiology conference*, IEEE, 2016, pp. 1033-1036.

[12]. A. T. Balaei, K. Sutherland, P. A. Cistulli and P. de Chazal, "Automatic detection of obstructive sleep apnea using facial images", *Biomedical imaging (ISBI 2017), IEEE 14th International symposium on*, 2017, pp. 215-218.

[13]. Haidar R, Koprinska I, Jeffries B, "Sleep apnea event prediction using convolutional neural networks and markov chains", *International joint conference on neural networks(IJCNN)*,doi:10.1109/ijcnn48605.2020.9207345.

[14]. Haidar R, Koprinska I, and Jeffries B, 2017 "Sleep apnea event detection from nasal airflow using convolutional neural networks", *International conference on neural information processing*, Guangzhou, 2017, pp. 819-827.

[15]. Syed M. S. Islam, Hassan Mahmood, Adel Ali Al-Jumaily, Scott Claxton, 2018 "Deep Learning of Facial Depth Maps for Obstructive Sleep Apnea Prediction", in *2018 International conference on machine learning and data engineering (ICMLDE)*, San Francisco, USA, 23-25 October, 2018, pp. 154-157.

[16]. AR Face Database: Yale face dataset-http://vision.ucsd.edu/content/yale-face-database.

[17]. K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *CoRR*, vol. abs/1512.03385, 2015

[18]. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, "Rethinking the inception architecture for computer vision", *IEEE Conference on computer vision and pattern recognition (CVPR)*, pp. 2818-2826, 2016, doi: 10.1109/CVPR.2016.308.

[19]. Kalyanapu Srinivas, B. R. S. Reddy, "Modified Kernel Based Fuzzy Clustering for MR Brain Image Segmentation using Deep Learning", International Journal of Engineering and Advanced Technology, vol.8, pp. 2249-8958, August 2019.

[20]. Bala Brahmeswara. Kadaru et al. "A Novel Ensemble decision tree classifier using hybrid feature selection measures for Parkinson's disease prediction", Int. J. Data Science (IJDS), ISSN: 2053-082X, Vol.3, No.4, 2018.

 DOI : 10.35291/2454-9150.2021.0288