

**ACKNOWLEDGEMENT**

The completion of any inter-disciplinary project depends upon cooperation, co-ordination and combined efforts of several sources of knowledge. We are grateful to **Mr. Ashish Tripathi** for his even willingness to give us valuable advice and direction whenever we approached her with a problem. We are thankful to her for providing us with immense guidance for this project.

We are also thankful to Dr. Chetna, Mr. Rupesh and Ms. Ruby for giving their valuable time for evaluating our project. We would also like to thank our College authorities and Head/Dean for giving us the opportunity to pursue our project in this field and helping us to successfully complete this project.

**Signature(s) of Students**

Anubhav Johri (9915103111)

Shivangi(9915103131)

Himanshu Sharma (9915102080)

**ABSTRACT < CHANGE LANGUAGE A BIT, MAKE IT GENERALISED, DON’T MAKE IT LOOK LIKE IMAGE BASED DETECTION>**

An automated detection system for Parkinson’s disease (PD) , employing the Deep Learning model i.e., Convolutional Neural Network (CNN) is proposed in this study. PD is characterized by the gradual degradation of motor function in the brain. Since it is related to the brain abnormality, electroencephalogram (EEG) signals are usually considered for the early diagnosis. In this work, the data from Parkinson’s Progression Markers Initiative (PPMI) database was obtained. PPMI is an observational, multicentre study that collects clinical and imaging data and biologic samples from various cohorts that can be used by researchers to establish markers of disease progression in PD . A thirteen-layer CNN architecture which can overcome the need for the conventional feature representation stages is implemented . In this project, we proposed Neural Networks framework which aims to help doctors and people in diagonising Parkinson Disease in early stages. Convolutional Neural Networks (CNNs) have been established as a powerful class of models for image recognition problems. Encouraged by the results, we will try to provide an extensive empirical evaluation of CNNs on large-scale image classification.

From a practical standpoint, there is very less amount of work on Parkinson Diagnosis based on Deep Learning. We have tried to create a system that will help in automatically diaganosing Parkinson Disease by classifying FP-SPECT images of normal and diseased subjects . This deep learning model pre-processes and divides the images into a number of frames with a fixed size. These frames are then tagged and classified by each node in every layer of the network. The frames are also compared with each other to find the accuracy percentage.

**TABLE OF CONTENTS**

**Topic Page No.**

*Acknowledgement i*

*Abstract ii*

*Table of contents iii*

*List of Tables iv*

*List of Figures iv*

*List of Abbreviations iv*

**Chapter 1: INTRODUCTION 1**

* 1. MOTIVATION BEHIND THE CHOICE OF TOPIC 1

1.2PURPOSE OF PROJECT 1

1.3 DESCRIPTION OF PROJECT 1

**Chapter 2: LITERATURE SURVEY 2**

2.1 INTEGRATED SUMMARY OF LITERATURE SURVEY 6

2.2 CURRENT/OPEN PROBLEMS 8

2.3 PROBLEM STATEMENT 9

2.4 OVERVIEW OF PROPOSED SOLUTION 10

2.5 TASK DIVISION 11

**Chapter 3: REQUIRMENT ANALYSIS 12**

3.1SOFTWARE REQUIREMENTS 12

3.2HARDWARE COMPONENTS 12

3.3FUNCTIONAL REQUIREMENTS 12

3.4NON-FUNCTIONAL REQUIREMENTS 12

3.4USER REQUIREMENTS 12

**Chapter 4: ANALYSIS, DESIGN AND MODELING 13**

4.1 OVERALL ARCHITECTURE 13

4.2 PROPOSED ALGORITHM 15

4.3 RISK ANALYSIS AND MITIGATION PLAN 16

4.4 TEST PLAN 16

4.5 IMPLEMENTATION 17

**CONCLUSION AND FUTURE SCOPE 20**

**APPENDIX 21**

REFERENCES 21

**LIST OF TABLES**

**Table Title Page No.**

2.1 Integrated summary of literature survey 6

4.3 Risk Analysis and Mitigation Plan 16

**LIST OF FIGURES**

**Figure Title Page No.**

2.1 Task division 11

4.1 Flow chart 14

4.2 Activity diagram 14

4.3 Code of dividing a video into frames in a fixed time interval 17

4.4 Set of frames obtained 17

4.5 Closer view of frame6 18

4.6 Closer view of frame46 18

4.7 Closer view of frame68 19

4.8 Closer view of frame90 19

**ABBREVIATIONS**

1. IEEE - Institute of Electrical and Electronics Engineers
2. CNN - Convolution Neural Network
3. RNN - Recurrent Neural Network
4. MPEG- Moving Pictures Experts Group
5. ANMRR - Average Normalized Modified Retrieval Rank
6. MPQF - MPEG Query Format

**Chapter 1: INTRODUCTION**

Parkinson Disease is a neurological disorder , which due to it’s low sensitivity of symptoms in early stages is very difficult to diagnose . So our study proposes a deep learning technique to predict PD in early stages.

**1.1 MOTIVATION**

Parkinson's disease (PD) is the second most common, long-term neurodegenerative disorder

of the central nervous system. In 2015, PD affected 6.2 million people and resulted in about

117,400 deaths globally. The costs of PD to society are high, the annual cost in the UK is estimated to be between £49 million and £3.3 billion, while the cost per patient per year in the U.S. is probably around US$10,000 and the total burden around US$23 billion. These facts alarmed a need for a cheap, efficient and an accurate way to diagnose PD in early stages so that a timely treatment could cure the patients before the disease becomes uncurable . With the help of our study we want to innovate a Deep Learning model which could save a million of lives and millions of money spent on PD throughout the world. We want to contribute to society by helping people suffering from PD and decreasing the number of deaths caused by PD every year.

**1.2 PURPOSE OF PROJECT**

Parkinson's disease is a progressive nervous system disorder that affects movement. Symptoms start gradually, and takes years to be noticeable . Hence , it becomes very difficult to diagnose PD at an early stage. PD can be cured if taken care of in an early stage , but mostly results in death as the time progresses. This study aims at diagnosing it by using Deep Learning technique CNN(Convulational Neural Networks) at an early stage hence increasing the probability of the patient getting cured .

**1.3 DESCRIPTION OF PROJECT**

The project provides an insight about all the decisive features like biomarkers from Cerebrospinal Fluid measurement (CSF) , dopamine transporter imaging, and FP-SPECT imaging to predict the Parkinson Disease and takes these features as an input to our Neural Networks in order to predict the disease. It also presented a critical analysis of the existing technologies which can be integrated together to achieve the goal of Automatic Parkinson Disease Detection.

**Chapter 2: LITERATURE SURVEY**

Research Paper [1] represents the results to evaluate the effectiveness of MPEG7 Color descriptorsin Visual Surveillance retrieval problems. Color descriptors from the MPEG7 standard are used,including Dominant Color, Color Layout, Color Structure and Scalable Color. Experiments are presented that compare the performance of these, and also compare automatic and manual techniques to examine the sensitivity of the retrieval rate on segmentation accuracy. In addition, results are presented on innovative methods to combine the output from different descriptors, and also different components of the observed people. The evaluation measure used is the ANMRR, a standard in Content-Based Retrieval experiments.

In the research paper [2] they conducted an empirical evaluation of MPEG-7 visual part of experimentation model (XM) color descriptors in a challenging problem of content-based retrieval of semantic image categories. The performance of the four color descriptors provided in the current XM reference implementation, Color Layout, Color Structure, Dominant Color and Scalable Color, is compared to that of HSV autocorrelogram, which has done well in recent empirical studies. Experimental results show that Color Structure provides best retrieval accuracy, whereas the computationally most expensive descriptor, Dominant Color, is worst in this problem.Considering the difficulty of the problem, Color Structure achieves a decent 35% average precision in retrieving the first 10% of the images from the semantic category of the query image. Dominant Color, which is the most expensive XM Descriptor in computational terms, provides the worst retrieval performance.

(Auto)correlograms capture both global occurrence statistics and local spatial organization of colors by a simple spatial constraint. Future work includes validation whether performance could be improved by a more extensive, possibly color independent, spatial constraint. Also, the de facto interpretations of (auto)correlograms are feature vectors, which largely for computational reasons are compared with L1 norm. However, (auto)correlograms are essentially histograms, and better performance could be achieved with information theoretic similarity measures, albeitat higher computational cost.

In the above research papers we see that earlier this classification of videos was done manually with the help of MGEP-7 descriptor. In the coming research papers we see how this classification is done automatically with the help of deep learning, which is then applied in our project.

In the research paper [3],it is told that GPU-enabled hardware is needed for the first time for indexing a single pass of feature vector extraction and storing into the database automatically. They have shown in this work that feature vector fv belongs to ℛ^1024 extracted by CNN [6] contains enough semantic information for segmenting raw video into shots with 0.92 precision; retrieving video shots by keywords with 0.84 precision; retrieving videos by sample video clip with 0.86 precision and retrieving videos by online learning with 0.64 precision.

They have shown in this work that all query types i.e., QueryByMedia, QueryByFreeText,

SpatialQuery and TemporalQuery can be implemented using the semantic features extracted from video by deep learning algorithms, namely by convolutional neural networks. Their contribution is presenting a video indexing and retrieval architecture based on unified semantic features and capable to implement MPQF query interface and sharing the results of real-world testing.

In Research paper [4], they have reviewed two lines of research aiming to stimulate the comprehension of videos with deep learning: video classiﬁcation and video captioning.

While video classiﬁcation concentrates on automatically labelling video clips based on their semantic contents like human actions or complex events, video captioning attempts to generate a complete and natural sentence, enriching video classiﬁcation’s single label to capture the most informative dynamics in videos. They have reviewed basic deep learning modules like CNN(Convolution Neural Network), RNN(Recurrent Neural Network) that have been widely adopted in the literature for video analysis. Long short-term memory (LSTM) is reviewed which overcome cons of RNN like —vanishing and exploding gradients. It is a variant that was designed to store and access information in a long time sequence. For Video Captioning, they have tried Supervised Deep Learning and Unsupervised Deep Learning on different datasets.

In the research paper [5], they have proposed the construction of a YouTube Recommender Network (YRN) and a recommender system derived from it. The YRN was created from the data collected from the YouTube website using their API. The data was composed of a number of videos and comments on each video. The YRN is undirected and weighted. The nodes represent the videos, whereas the edges are established between two nodes if there is a user who commented on both of them. The edge weight represents the number of times two nodes are associated with comments from different users in both of them. After the YRN was generated, they observed scale-free and small-world properties in the network with a number of communities. It was demonstrated that the distribution of tags inside communities is diverse and follows a power law. The weight of an edge predicts the strength of the relation between the two nodes it connects. A utility value was also introduced which represents the importance of a node. The higher the utility value, the more important the node is. Finally, a way to build a recommender system derived from our YRN was proposed in which firstly the videos are recommended to users from the highest utility value to the lowest. Secondly, when a user is watching a video, other nodes connected to it were recommended.

Similarly, the research paper [6] has introduced ViTS, an industrial Video Tagging System which generates tags based on information crawled from the Internet and learns relations between concepts. The core of the system is a knowledge base that is constantly updated to capture the dynamics of the indexed concepts. ViTS was tested on a subset of videos from the YouTube- 8M dataset. The tags generated by ViTS were highly graded by human users exposed to a visual summary of the video and its metadata. The accuracy of 80.87% is comparable to the inter-annotator agreement of (non-expert) humans in the task of semantic annotation. This high quality, combined with its capability of capturing not-only visual concepts, shows the capability of ViTS as a rich video indexing system. Moreover, experiment results on Youtube-8M are publicly available. The presented tagging system shows how contextual data is a powerful source of information when indexing web videos. Exploiting the relations between concepts allows generating a rich set of tags with a light computation, desirable when addressing a web-scale indexing. However, content-based techniques could also extend our content based tags. Our future work will address exploiting these tags as weak labels for computer vision and audio processing deep models, which have been shown impressive recognition performances in the recent years.

Research Paper [7] studies the exploration of user searching behavior through click-through data, which is largely available and freely accessible by search engines, for learning video relationship and applying the relationship for the economic way of annotating online videos. They have demonstrated that, by a simple approach using co-click statistics, promising results were obtained in contrast to feature-based similarity measurement.

A new method based on polynomial semantic indexing is proposed to learn a latent space for alleviating the sparsity problem of click-through data. The proposed approaches are then applied for three major tasks in tagging: tag assignment, ranking, and enrichment. On a bipartite graph constructed from click-through data with over 15 million queries and 20 million video URL clicks, they have shown that annotation can be performed for free with competitive performance and minimum computing resource.

In the research paper [8] they have explained what is tagging and geo-tagging. It mentioned that there are still a number of videos that are untagged and the need to do so in order to improve the performance of multimedia retrieval. The paper is based on the three techniques of media eval 2010 benchmark initiative which are discussed later in the paper. For auto-tagging in a video, two approaches were used, first is called extraction which tags videos on the bases of its metadata. The second approach is called assignment in which tags are assigned from a fixed data set. Auto spoken audio tagging was also applied but was not possible due to noise in the voice. It was only possible on speech on telephones. For geo-tagging spatial distribution was used in which location is represented as grid cells. The three techniques of media eval 2010 benchmark initiative are then discussed which are: tagging task professional which needs human activity for assigning tags from a closed tags using mean average precision(MAP) technique, tagging task wild wild web which includes tagging by users in an online video community using MAP and metadata, finally is placing task which requires participants to automatically assign geo-ordinates to the video.

Automatic video tagging was also done in research paper [9] which tells how video content on the World Wide Web continues to expand and it is considerable to annotate videos for effective and accurate search and mining properly. The paper says that while the idea of annotating imagery using keywords is simple and well known, it facilitates basic for annotating videos with natural keywords automatically to enhance search is a significant arising problem with the excessive potential to improve the quality of video search. The paper then discusses the benefits of levelling large scale video datasets for automated annotation also presents fresh challenges and requires methods specialized for scalability and efficiency.

**2.1 INTEGRATED SUMMARY OF LITERATURE SURVEY**

|  |  |  |
| --- | --- | --- |
| Research paper name | Author | Summary |
| “A deep learning approach for Parkinson’s disease diagnosis from EEG  Signals” | Shu Lih Oh Yuki ,  Hagiwara,  U.Raghavendra,  Rajamanickam Yuvaraj,  N. Arunkumar,  M.Murugappan,  U. Rajendra Acharya | Electroencephalogram (EEG) signals are considered as an abnormality to detect PD.A thirteen-layer CNN architecture which can overcome the need for the conventional feature representation stages is implemented. The developed model has achieved a promising performance of 88.25% accuracy, 84.71% sensitivity, and 91.77% specificity. |
| “Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging” | Hongyoon Choia, Seunggyun Haa,  Hyung Jun Ima,  Sun Ha Paekd, Dong Soo Lee | Dopamine transporter (DAT) imaging such as123Ifluoropropylcarbomethoxyphenytropane (FP-CIT) single-photon emission computed tomography (SPECT) is used as one of the tools for the diagnosis of Parkinson's disease (PD). Study aimed to develop an automated FP-CIT SPECT interpretation system based on deep learning for the objective diagnosis. Input data were SPECT images downloaded from the PPMI database. |
| “An Improved Approach for Prediction of Parkinson’s Disease using Machine Learning Techniques” | Kamal Nayan Reddy Challa,  Venkata Sasank Pagolu,  Ganapati Panda,  Babita Majhi | In this paper non-motor features such as RBD and olfactory loss are used.Research has used automated diagnostic models using Multilayer Perceptron, BayesNet, Random Forest and Boosted Logistic Regression. It has been observed that Boosted Logistic Regression provides the best performance with an impressive accuracy of 97.159 % and the area under the ROC curve was 98.9%. |
| “Analysis and Identification of Parkinson Disease based on fMRI” | Anshul Singh , Shaurya Singh, Vijay Khare, Neha Mehra, Dr Shamim Akther ,  Dr Chakresh Kumar Jain | This study has used the neuroimaging techniques *viz* fMRI and invasive electroencephalography methods that are applied for the diagnosis of disease state. Functional magnetic resonance imaging (fMRI), is a non invasive method, has proven useful for studying the behavior (neural activity) of human brain through functional brain network analysis of the generated images. |
| A recommender system for youtube based on its network of reviewers.” Social computing (socialcom), 2010 ieee second international conference on. IEEE, 2010. | Qin, Song, Ronaldo Menezes, and Marius Silaghi | They have proposed the construction of a YouTube Recommender Network (YRN). It was created from the data collected from the YouTube website using their API.  A recommender system derived from YRN was proposed in which firstly the videos are recommended to users from the highest utility value to the lowest. Secondly, when a user is watching a video, other nodes connected to it were recommended. |
| Video tagging system from massive web multimedia collections.” Proceedings of the 5th Workshop on Web-scale Vision and Social Media (VSM). IEEE Press, 2017. | Fernández, Delia, et al. “ViTS: | This has introduced ViTS, an industrial Video Tagging System which generates tags based on information crawled from the Internet and learns relations between concepts. The core of the system is a knowledge base that is constantly updated to capture the dynamics of the indexed concepts. The accuracy obtained is 80.87%. |
| “Automatic tagging and geotagging in video collections and communities.” Proceedings of the 1st ACM international conference on multimedia retrieval. ACM, 2011. | Larson, Martha, et al. | In this, they have explained what is tagging and geo-tagging. The paper is based on the three techniques of media eval 2010 benchmark initiative: tagging task professional which needs human activity for assigning tags from a closed tags using mean average precision (MAP) technique, tagging task wild wild web which includes tagging by users in an online video community using MAP and metadata, finally is placing task which requires participants to automatically assign geo-ordinates to the video. |
| “Semantic video search by automatic video annotation using TensorFlow.” Manufacturing & Industrial Engineering Symposium (MIES). IEEE, 2016. | Ashangani, Kithmi, et al. | Automatic video tagging was also done in research paper [8] which tells how video content on the World Wide Web continues to expand and it is considerable to annotate videos for effective and accurate search and mining properly. It facilitates basic for annotating videos with natural keywords automatically to enhance the quality of video search. |
| Ojala, Timo, Markus Aittola, and EsaMatinmikko. "Empirical evaluation of MPEG-7 XM color descriptors in content-based retrieval of semantic image categories." Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 2. IEEE, 2002. | Timo Ojala, Markus Aittola and EsaMatinmikko | The performance of the four color descriptors:- Color Layout, Color Structure, Dominant Color and Scalable Color, Experimental results show that Color Structure provides best retrieval accuracy, whereas the computationally most expensive descriptor, Dominant Color, is worst in this problem. |

Table 2.1 Integrated summary of literature survey

**2.2 CURRENT PROBLEMS <TO-DO>**

Given the substantial amounts of videos generated at an astounding speed every hour and every day, it remains a challenging open problem how to derive better video representations with new methodologies, the abundant interactions of objects and their evolution over time with limited supervisory signals to facilitate video content understanding (i.e. , the recognition of human activities and events as well as the generation of free-form and open-vocabulary sentences for describing videos)

In doing Video Tagging using Mining User Search behavior, there is a problem with partial visual near duplicates, which frequently happen in Web videos. The visual features can be exploited together with click through and document features for a more comprehensive manner of characterizing visual similarities. Similarly, in Deep Learning Methodology like CNN, performance of sample-based video retrieval should be enhanced. Several approaches should be explored for lowering the feature vector dimensionality in order to search in log time scale, e.g. random projections and compact binary descriptors.

While using MPEG-7 Descriptor, again some problem arises like it fails to adequately address certain issues familiar to Visual Surveillance researchers. For example, the rank ordering method cannot in itself provide evidence that a given query example does not appear in a dataset: there will always be one element of the data set most similar to the example. Similarly, probabilistic estimates of identity, for incorporation with other uncertain cues, are not easily deduced from the rank method. One challenge is the uniﬁcation of retrieval metrics across the research communities. Also, Multi-camera retrieval rate should be enhanced by specification of a pre-processing method which is lacking in color constancy.Color Structure achieves a decent 35% average precision in retrieving the first 10% of the images from the semantic category of the query image. Dominant Color, which is the most expensive XM Descriptor in computational terms, provides the worst retrieval performance. This is understandable, since the descriptor is targeted for presenting local features such as regions or objects, not complete images.

**2.3 PROBLEM STATEMENT**

To efficiently develop a model , which predicts whether a person is suffering from Parkinson Disease or not and compare it’s results, performances with the state-of-the art work.

**2.4 OVERVIEW OF PROPOSED SOLUTION**

Deep Learning techniques are widely used for disease detection in various medical researches because it excels in nonlinear feature extraction and combines both feature extraction and classification processes. Hence we propose Deep Learning technique CNN as a solution, with different features like Cerebrospinal Fluid measurement (CSF) , dopamine transporter imaging, FP-SPECT imaging , and olfactory loss decided over the course of time.

So, to increase efficiency as well as accuracy of our system, we need to modify our technique.

**2.5 TASK DIVISION <TO DO>**

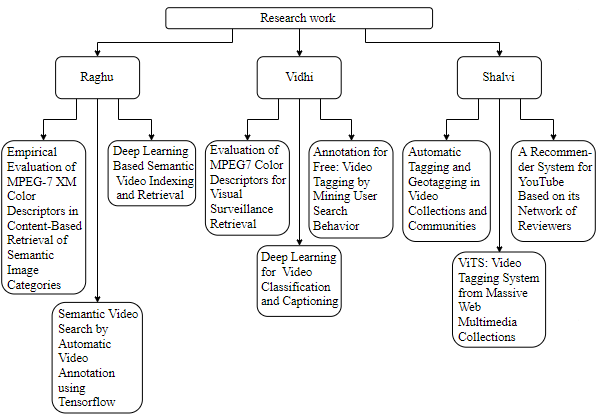


Fig. 2.1 Task division

**Chapter 3: REQUIREMENT ANALYSIS**

**3.1 SOFTWARE REQUIREMENTS**

* Windows 7 and above , or Ubuntu or Centos
* Python Idle, Anaconda

**3.2 HARDWARE COMPONENTS**

* Processor – i5
* Memory – 4GB RAM

**3.3FUNCTIONAL REQUIREMENTS**

* The user should have information of the libraries used.
* The user should be familiar with python.
* The user should be familiar with Anaconda.

**3.4NON-FUNCTIONAL REQUIREMENTS**

* The program should not have any reliability issues. The program will be thoroughly tested.
* The program should run on any software mentioned above.
* The program should be able to classify images,sounds.

**3.5USER REQUIREMENTS**

* User should be able to work in python.
* User should be able to work in Anaconda.

**Chapter 4: ANALYSIS, DESIGN AND MODELING**

**4.1 OVERALL ARCHITECTURE**

The previous sections presented a critical analysis of the existing technologies which can be integrated together to achieve the goal of Parkinson Disease Detection. In this section, new design and architectural proposals will be made for the target system.

Our dataset had been created using PPMI database which can be accessed using “www.ppmi.info”.<COMPLETE IF FEELS NECESSARY OR DELETE THE WHOLE CHAPTER>

**4.3 RISK ANALYSIS AND MITIGATION PLAN**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Risk ID | Description of Risk | Risk Area | Probability (P) | Impact (I) | RE = PxI | Risk Selected for Mitigation on | Mitigation on plan | Contingency plan |
| 1 | It requires system with effective GPU | Hardware | 3 | 3 | 9 | N | N | Use Google Colab |
| 2 | Large dataset is required | Algorithm | 1 | 5 | 5 | Y | Y | Browse PPMI ,IDA sites |

Table 4.3 Risk Analysis and Mitigation Plan

**4.5 IMPLEMENTATION**

Fig. 4.3 Code of dividing a video into frames in a fixed time interval

Fig. 4.4 Set of frames obtained

**CONCLUSION AND FUTURE SCOPE**

To come up with an efficient way detect Parkinson Disease , we had studied different research papers based on Parkinson Detection using various Machine Learning and Deep Learning algorithms to classify subjects into the category of normal and suspicious. We compared results obtained by various studies using different techniques and came at a conclusion of implementing Deep Learning for our research.

In future, we will try to make the detection more efficient and results more accurate by adding more important decisive features like olfactory sound and handwriting-check with Deep Learning methodologies to overcome drawbacks of previously researched work/project . Also, we will look into new algorithms that could reduce the computation needed and would also make the system lighter.

**APPENDIX**

**REFERENCES**

[1] Annesley, James, James Orwell, and J-P. Renno. "Evaluation of MPEG7 color descriptors for visual surveillance retrieval." Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on. IEEE, 2005

[2]Ojala, Timo, Markus Aittola, and Esa Matinmikko. "Empirical evaluation of MPEG-7 XM color descriptors in content-based retrieval of semantic image categories." Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 2. IEEE, 2002.

[3]Podlesnaya, Anna, and Sergey Podlesnyy. "Deep learning based semantic video indexing and retrieval." Proceedings of SAI Intelligent Systems Conference. Springer, Cham, 2016.

[4] Wu, Zuxuan, et al. "Deep learning for video classification and captioning." arXiv preprint arXiv:1609.06782 (2016).

[5]Qin, Song, Ronaldo Menezes, and Marius Silaghi. "A recommender system for youtube based on its network of reviewers." Social computing (socialcom), 2010 ieee second international conference on. IEEE, 2010.

[6] Fernández, Delia, et al. "ViTS: Video tagging system from massive web multimedia collections." Proceedings of the 5th Workshop on Web-scale Vision and Social Media (VSM). IEEE Press, 2017.

[7] Ting Yao †, Tao Mei ‡, Chong-Wah Ngo †, Shipeng Li Annotation for Free: Video Tagging by Mining User Search Behavior, IEEE 04-September 2017

[8] Larson, Martha, et al. "Automatic tagging and geotagging in video collections and communities." Proceedings of the 1st ACM international conference on multimedia retrieval. ACM, 2011.

[9] Ashangani, Kithmi, et al. "Semantic video search by automatic video annotation using TensorFlow." Manufacturing & Industrial Engineering Symposium (MIES). IEEE, 2016.