# VIDEO POPULARITY PREDICTION USING STACKED BILSTM LAYERS

***Neeti Sangwan[1*], Vishal Bhatnagar[2]***

[1]Research Scholar, GGS Indraprastha University, India

[2]Professor, NSUT East Campus (Formerly Ambedkar Institute of Advanced Communication Technologies and Research), New Delhi, India

[1]Assistant Professor, MSIT, New Delhi, India

Email: neetisangwan@gmail.com[1*](corresponding author), vishalbhatnagar@yahoo.com[2]

DOI: https://doi.org/10.22452/mjcs.vol34no3.2

**ABSTRACT**

*Social media is now not only limited to being a life event sharing platform, but it also has evolved as a monetary medium. Advertisements showing on popular videos may result in more sales conversion. So it is of utmost interest to predict the popularity of videos before uploading it on the platform. In this research article, we propose a deep learning algorithm to predict the popularity of YouTube videos. With the content and temporal features of the YouTube videos dataset, we use a novel stack of deep learning layers. We validate the approach with state-of-the-art methods and prove that the proposed complex stacked architecture gives more accurate and stable results. Results are also tested for short duration prediction with a different number of reference days after video publishing.*

**Keywords:** ***Videos, Prediction, Deep Learning, Regression***

## 1.0 INTRODUCTION

The internet is an essential requirement for every human being. Tons of users are available on different social media websites like Twitter, Facebook, YouTube, etc. Online video streaming plays a vital role in the mobile Internet. The user behavior or interest can be recognized by the user's interest in videos over the internet. The number of users on YouTube and Netflix has become thrice since 2013 [1]. The shared video's characteristics define the nature of the creator and the type of viewers that explore their user behavior. The content of the video is the key factor for its popularity. The popularity prediction of online video is provided with the benefits of public management and security concerns [2]. As per the business point of view, the popularity prediction analyses the latest trend and innovations globally. The most popular video platform, YouTube, helps in revenue generation on behalf of the user's watch time for a video. More content-rich is the video, the more traffic and popularity it will attain. The popularity prediction of online streaming video is a difficult task due to the large number of users available on social media, and it creates multidimensional content every hour.

The popularity prediction can help the users to monetize with their videos as well as for the video-sharing platforms too. The impact of commercial advertisements can be maximized during popular programs, which increases their business [3]. The configurations of the broadcast network are optimized in advance if popularity prediction is performed earlier. Some social media platforms like Vine and Snapchat generate short time videos of approximately 6 to 10 seconds. However, YouTube videos are more extended versions, and their characteristics like audio, face, background, etc. are analyzed very efficiently. The social media popularity prediction efficiently improved the promotion of any video on that platform [4]. It also helps in creating the video related to the same content. Several schemes are used for the prediction of video popularity on social media content. The prediction model schemes are classified into three categories; 1) Generative Model, 2) Time-Series, and 3) Feature-based approach. The generative model, like Hawkes, is used for the prediction of Twitter content. It recognizes the intrinsic characteristics of the tweet from time to time [5]. The generative method is to deal with the stochastic data available on the social media network. These are not used for the YouTube video popularity prediction due to complex content propagation [6].

The time series based approaches use the regression model for the prediction popularity of video content. The regression models utilized in state of the art are Auto-Regressive and Moving Average (ARMA) [7], Follow-The-Regularized Leader (FTRL) [1], Random Forest (RF) [3], and Transductive Low-Rank Multi-View Regression (TLRMVR) [8]. The early-stage prediction is not provided in Time series methods. The prediction is based on the historical data popularity, so cold start problems turned out in these prediction methods [6]. The third category of the popularity prediction method is a feature-based model — some deep learning models are also used for feature extraction from the online video contents. The features like user information, user interest, video contents, and time of sharing by all users are extracted.

The contribution of the work is:

- This research work extends the previous work on YouTube video's popularity prediction. We propose a novel stacked deep learning architecture to predict the short-term forecast.
- Our contribution is in designing a complex deep network that can abstract the input data to depth. This complex network uses Bidirectional Long Short-Term Memory (biLSTM) layers since Long Short-Term Memory (LSTM) has shown its strength in similar prediction works.
- Our work utilizes the YouTube videos' temporal and statistical features to feed into the network.

Further, in the paper, section 2.0 analyzes the previous researchers' efforts. Section 3.0 extensively discusses the YouTube data for the experiment and we present the proposed deep learning model in section 4.0. Section 5.0 discusses the state of art comparisons with our network outcome and work is concluded in section 6.0.

## 2.0 BACKGROUND

Huang et al. proposed the ARMA model for the popularity prediction of real-time scaling online videos based on their temporal features and the interest of users [7]. ARMA least frequently used model outer perform than the least recently perform while caching the video. In [2], an in-depth learning approach was utilized to predict text content available on social media. The neural network-based predictive model extracts the features and recognizes popularity as per user interest or reactions of real-time data sets. A recent hybrid technique was developed by combing the XG Boosting and deep learning for the popularity prediction of Metadata content. The volatility of video content was resolved by FTRL proximal algorithm. Zhu et al. proposed the Dynamic Time Warping (DTW) distance-based K-medoids algorithm for extracting the four trends among the group of videos, then the Random Forest (RF) regression model was used to build the four specific predictive models. The experiment was tested on the broad set of real Video on Demand (VoD) data from Jiangsu Broadcasting Corporation and configures better prediction performance than the existing algorithm [3].

In [5], a generative approach Hawkes was proposed for the prediction popularity of online social media platform-Twitter. A generative model was implemented to predict the text popularity of the Twitter comment dataset. The relative error is minimized among the tweets for 10 minutes. It has both generative and feature model properties that provide higher prediction accuracy. The TLRMVR was proposed for the popularity prediction of short-time videos [8]. The prediction performance of proposed TLRMVR is improved by jointly considering the intrinsic representation of source and target samples. A multi-graph regularization approach is also added to improve the generalization capability and prevent over-fitting problems. The interpretability of end-to-end popularity prediction methods is enhanced by weighted Grad-Cam results of sequential inputs with self-attention schemes [4]. The Spatio-temporal heatmaps are created which indicated the video frames and contributes to the video popularity prediction of online video creators.

A latent vector represents the knowledge base (KB) entity and compactly encoded the KB information to improve the popularity prediction of online video. LSTM model was integrated to enhance the dynamic performance of the KB entity based predictive model [9]. The author reflects [6] fast prediction of YouTube video contents using the leveraging history of that video. A Lifetime Aware Regression Model (LARM) model was proposed to predict popularity using the time-to-time content of videos. The performance is evaluated with a YouTube dataset for different period intervals. The Fine-grained Video Attractiveness Dataset (FVAD) of popular TV serials was tested via deep learning attractive models [10]. The famous episode of TV serial was extracted based on their visual and audio characteristics. The correlation among the episodes and nine user engagement behaviors was also tested using a deep learning algorithm. A PreNets unifies two thinking paradigms: features driven and point process methods in an adversarial manner. The features-driven model predicts popularity based on the spatial information, temporal information and point process to recognize the dynamic pattern for evaluating the popularity of online data content. The training loss of both methods was minimized by Wasserstein learning based two-player minimax game model. The proposed PreNets attained significant improvement in Mean Absolute Percent Error (MAPE) for Twitter Cascades and Amazon review prediction [11]. The stability of learning can be improved by the Wasserstein Generative Adversarial Networks (WGAN) [12]. Researchers have utilized the variants of WGAN in medical [13], speech processing [14], security [15], image processing [16] and other distinct fields.

A support vector regression model with Gaussian radial basis functions was proposed in [17] for the popularity prediction of online video contents. The experiment was tested on YouTube and Facebook datasets, which contain 2400 videos. The social and visual features were extracted from the videos and classified the popularity using Support Vector Machine. Mekouar et al. proposed the logistic regression method for the popularity prediction of YouTube videos. The stepwise regression method also implements on logistic regression, which improved the prediction accuracy [18]. The Hawkes Intensity process (HIP) was used in [19], which predicts the popularity of historical videos. It provides flexibility to the users and content producer related to the viral video. The virtual quality of video content was compared with the endo-exo map visualization. A series of applications were enabled in the case of HIP, which compares the videos and channels. The popularity of video was predicted based on which promotion of video will be performed in the future. The set of aggregate engagement metrics for online videos were measured in [20] with average watch time, average watch percentage, and relative engagement. The 5.3 million videos from publically available datasets tested on metrics configuration. This work was not translated to user-specific engagement. A time series classification using Long Short-Term Memory with the Fully Convolutional Network (LSTM-FCN) was utilized in [21], multivariate time series classification in [22] with LSTM-FCN and Attention Long Short-Term Memory with the Fully Convolutional Network (ALSTM-FCN) for activity recognition. The interconnection parameters of the YouTube dataset were analyzed by Imperative Structural Modeling (ISM) in the study [23]. The popularity of online video content was configured by interviewing their creator. The information related to the video making was shared, and popularity prediction was performed [24].

A proactive music video caching scheme was developed using Convolutional Neural Network (CNN) classification and multivariate popularity prediction of YouTube data. The CNN was used to derive the music features and as a mood classifier. The three policies: mood, genre, and popularity were evaluated in the study [25]. The video's future popularity was predicted using features of the videos and links related to the User Generated Content (UGC) videos in [26]. The YouTube UGC videos were tested via elements exploiting the approach. A new scheme, Social network assisted Video Prediction (SoVP), was proposed for the prediction popularity of online social video content. The internal content of the video and propagation structure was considered in the SoVP scheme [27]. A cross-domain real-time transfer approach based on Social transfer was proposed for the online video popularity prediction. The cross-domain process boosted the knowledge of various social media applications [28]. The tag enrichment by comment section on a social media video improved the prediction of popularity. The achievable scaling law depends on practical popularity distribution was used for the video device to devise allocation. The MZipf video popularity distribution was adopted for experimental setup, which measures the demand of the largest video service in the UK only for cellular users [29].

An analysis of BiliBili, a site having user-generated content with unique features like comment, chats, and virtual money donation, was done [30]. The popularity of UGC video content was predicted, characterized the video repository and activities of a user. Further graph model was presented for user relationships and high-level social structures [30]. A content and location feature based approach was proposed for content popularity prediction. The two algorithms: Ridge regression Prediction with Upper Confidence (RPUC) and $H_\infty$ filter Prediction with Dynamic Threshold (HPDT) were used for content popularity prediction, which deals with the location caching decisions [31]. A novel approach, Constant History Exponential Kernels and Social Signals (CHESS), was proposed for the popularity prediction of social media videos. CHESS is scalable due to constant space requirements per video and easy to handle the workload on Facebook. The CHESS works based on historical patterns, and other learning modal features, so accurate prediction is obtained [32].

A new scheme Deep Temporal Context Network (DTCN) was presented for sequential prediction of popularity using temporal context and temporal attention. The embedding of two collaborative networks was performed efficiently, and two types of secular contexts were applied for the popularity prediction of social media content. The sequential popularity prediction at a multiclass point was estimated by DTCN [33]. In [34], an EvoModel was developed for the dynamic popularity prediction of online social media content. It provided better popularity prediction of videos and explained the behavior of auxiliary factors controlling the video popularity dynamics. The early age technique was used in the EvoModel for the estimation of the parameters. Robert et al. proposed the engagement parameters for the popularity prediction of brand-related user posts. The experiment was tested on the fast-food brand related post available on the Instagram platform, which contains visual and textual features. More accurate popularity prediction was achieved using the engagement parameters like factual information, entertainment and sentiments [35]. A two-stage model was proposed for the online video future popularity prediction. The total number of views count was evaluated for the popularity prediction of online video. A large scale, long term, and updated database of online videos were formulated via the proposed two-stage method. In the first stage, the video features like statistical, textual and visual were extracted from the video; after that, the future view count prediction was performed [36]. The redundancy feature of distinct layers of CNN is exploited for compression of CNN to reduce the expenditure of computation and cost [37]. Neural Network Model for Multi-Aspect with Strong Correlation (NNMASC) was proposed to embellish the predictive recommendation [38].

A hybrid regression prediction model was implemented for the User Generated Video (UGV). The proposed model dynamically adapts the popularity of UGV using historical training datasets. The popularity prediction error was reduced by approximately 14% compared to the other offline and online popularity prediction schemes [39]. Tan et al. proposed the novel time series model for the popularity prediction of online videos. The time series models' basic principle depends upon the correlation among past and future popularity series. The most popular video was selected with the time series prediction instead of future view counts. The long term popularity prediction method was tested in the real world datasets [40].

## 3.0 DATASET DESCRIPTION

For YouTube video prediction, we required a dataset that has daily views/engagements recorded for many channels and videos. Unfortunately, YouTube API has discontinued many features which could help us to collect necessary attributes. We have to rely on the data collected in [20]. The author in [20] used YouTube API to collect various attributes like video ID, title, description, number of views, watch time, category duration, upload time, etc. Few of these features are discontinued now by YouTube. Therefore, we can categorize these attributes as:

- *Temporal features*: Attributes like video duration, upload time, video category, number of views, video quality, etc. are temporal features.
- *Engagement features*: Video popularity depends upon the engagement of the user with the video. Temporal characteristics may motivate the user to click on the video and watch the video; content should be engaging. The watch duration of the video is the only factor that conveys this information.

Temporal and content engagement features are made available in this data repository. This dataset has a significant advantage in that it has collected the daily watch time ($\omega_t$) of each video along with daily view counts. Daily Watch time is the primary engagement factor of video as a user may visit the video URL, but if the content of the video is not impressive, then it will drop out very soon. More is the number of dropping out users; less is the popularity. Watch time is higher for content-rich videos. The inconsistency in daily watch time and view time for video on automobiles is shown in Fig. 1. The view counts keep on decreasing after uploading, whereas the watch time doesn't follow a pattern. It has many perturbations, which may be due to other outer world factors that might have triggered the user's interest. View count shows a steady slope after 2 months but watches time is still high. It shows that the video must have good content. So, watch time has a key significance in predicting video engagement.

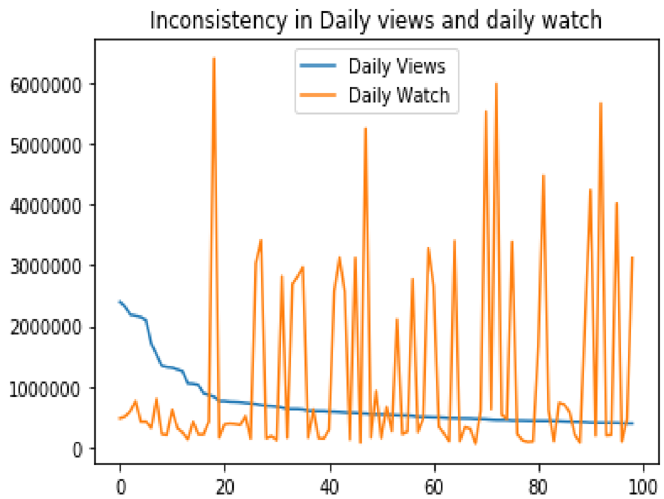



Fig. 1: Inconsistency in daily watch time and the view count of YouTube videos

The data available from this paper is in JSON format and has various categories of data. We serialized the time series data with respect to each day recording every video, and the data samples increased to days $\times$ n times. The average watch percentage and video duration Gaussian $pdf$ curve is plotted in Fig. 2 (a) for automobile videos. The block curve line is plotted for 50 percentiles of the data. 10, 30, 70, and 90 percentiles of watch duration are shown in blue color in Fig. 2 (a). It shows a linear decrease in slope with an increase in video duration. Though in some video categories, this insight might change. For example, a movie on YouTube might have a linearly increasing slope with watch duration. It leads to our assumption that the video duration is also another essential factor to influence the watch percentage. The shorter duration videos are seemingly more watched than longer videos. We have plotted a correlation matrix for the features in the dataset in Fig. 2 (b). The average watch percentage, watch, and view count is also plotted in Fig. 2 (b). The correlation between daily views and the daily watch is 0.55, whereas the average watch duration is correlated by 0.68, with average views for 30 days. It is also perceived from the correlation matrix that days, video quality (definition: HD or lower quality) doesn't affect the watch duration or views count. Significantly less correlation value is observed for these.

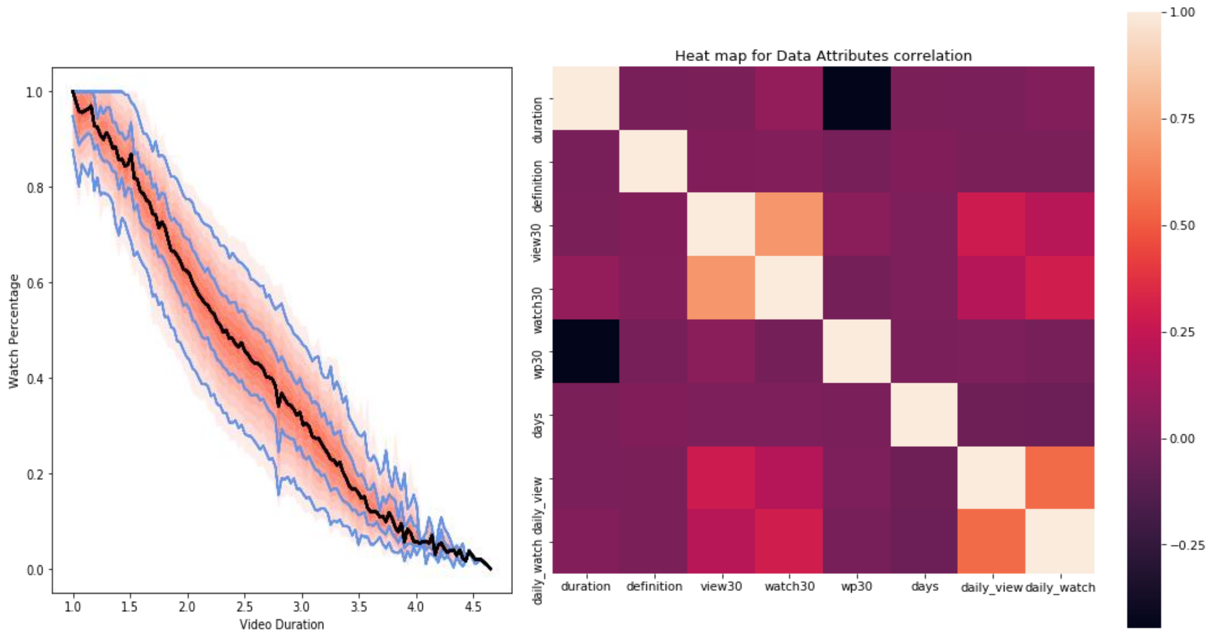



Fig. 2: (a) Pdf curve for average watch percentage v/s video duration (b) Correlation matrix amongst attributes in the data

## 4.0 PROPOSED POPULARITY PREDICTION SCHEME

In this paper, we cast the popularity prediction scheme as a time series prediction problem. The objective is to predict the popularity of YouTube videos at the time $t_t$, providing the video attributes upto the time $t_r$ (considerably $t_t > t_r$). The collected data is time-series data, and in the literature, the LSTM model is found very efficient to predict the time series problems. We propose to improve the prediction accuracy by a novel complex stacked deep learning model. Our model has the advantage of abstraction to the input data in depth. In our stacked network, we combined multiple biLSTM layers, and to the best of our knowledge, this arrangement of layers has not been published yet. In deep learning architecture, there is no single theory to design the network for any particular application. Every model is either suggested from the literature study or created after several hits and trials of several layers' combinations. In this work, too, we come up with the final model after several attempts. Though few researchers have also used heuristic optimization algorithms to optimize the layers' number, their filter/stride size, etc. but this method is very computationally expensive for a large dataset like ours. We started to test the architecture with a varying number of memory units in the LSTM network arrangement. We learn from the literature that biLSTM's forward and backward training is better than LSTM's either forward or backward training. The tentative number of memory units from the extensive testing is used in the final deep network with the first biLSTM layer. Multi-biLSTM layers will be used to fine filter the noise in the data and get a more accurate prediction. The proposed flow of the process is shown in Fig. 3. The data features are fed into the proposed stacked network without any pre-processing. Ahead in this section, we will discuss the LSTM architecture before introducing the proposed network.

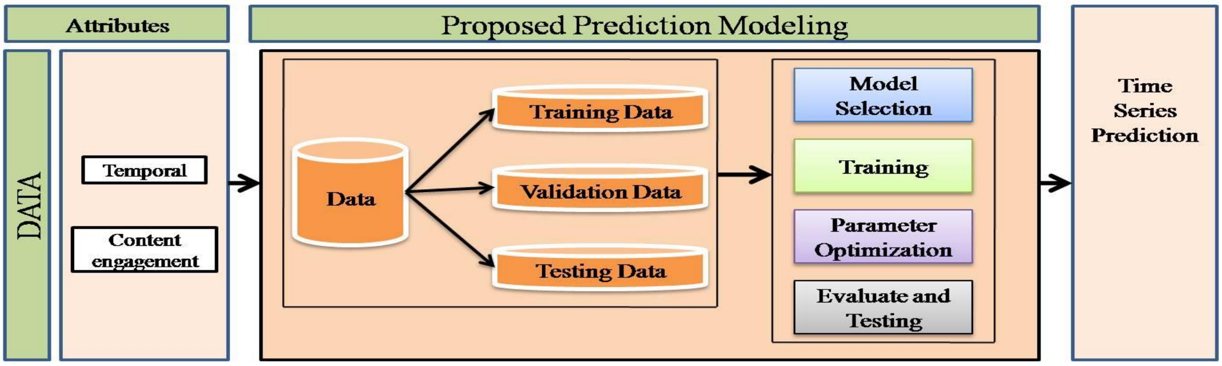



Fig. 3: Steps of proposed time series popularity prediction

### 4.1 Long Short Term Memory

The time sequence learning problems are solved by the Recurrent Neural Network (RNN). The long term dependencies cannot be captured by the RNN method. This limitation of RNN is overcome by the LSTM. The key feature of LSTM is that some gates that control the flow of information along the time axis. It can capture more accurate long term dependencies. The hidden layer of LSTM contains the memory blocks called cells. In each time step $t$, the hidden cell state $h_t$ is updated via fusion of data at same time step $x_t$, input gate, output gate, forget gate, update gate, and past hidden state value $h_{t-1}$. The structure of LSTM contains several cell state and gate functions in the hidden layer, as shown in Fig. 4. The input and output gate is present in each cell state or memory block. Both gates perform the control operation as input activation and output activation. Later forget gate is added to the cell state [31].

**Forget gate-** It is used to discard the information from the cell state. The sigmoid activation function provides control to the gate. The forget gate is represented by the $f_t$ described below;

$$f_t = \sigma(W_f.[h_{t-1}, X_t] + b_f \qquad (1)$$

Here $W_f$ shows the weight matrix in the forget gate, $h_{t-1}$ is the previous output, $X_t$ is a present input, $b_f$ is bias, and $\sigma$ is a sigmoid activation function.

**Input gate-** The key function of the input gate is to decide the updated value of new information in the cell state. It uses the tanh activation function, and the update gate provides control. The mathematical derivation of the input gate is

$$i_t = \sigma(W_i.[h_{t-1}, X_t] + b_i \qquad (2)$$

$i_t$ is an input gate notation, $W_i$ and $b_i$ are the weights and bias matrix of the input gate. The update gate provides control. The function of the update gate is to decide which latest information is saved in the memory blocks. It uses the sigmoid activation function and represented as

$$\hat{C}_t = \sigma(W_c.[h_{t-1}, X_t] + b_c) \qquad (3)$$

Here $W_c$ and $b_c$ are the weight and bias matrix of the update gate.

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \qquad (4)$$

In equation 4, $C_t$ and $C_{t-1}$ are the present and past cell state values.

**Output gate-** The output gate updates the final information of the cell state. The sigmoid activation function provides control to the output gate.

$$o_t = \sigma(W_o.[h_{t-1}, X_t] + b_o \qquad (5)$$

$$h_t = o_t \times \tanh(C_t) \qquad (6)$$

Here, $h_t$, $W_o$ and $b_o$ are the present output, weight, and bias matrix of output gate, respectively.

The LSTM model can only preserve information from the past as it considers the inputs only from the past. In the case of a biLSTM unit, the inputs run in two ways, one from the past to future (forward training) and the other from future to past (backward training). The combination of two hidden states can provide any point in time to preserve information from both the future and the past. The biLSTM model is better in terms of both forward/backward training than the unidirectional LSTM model.

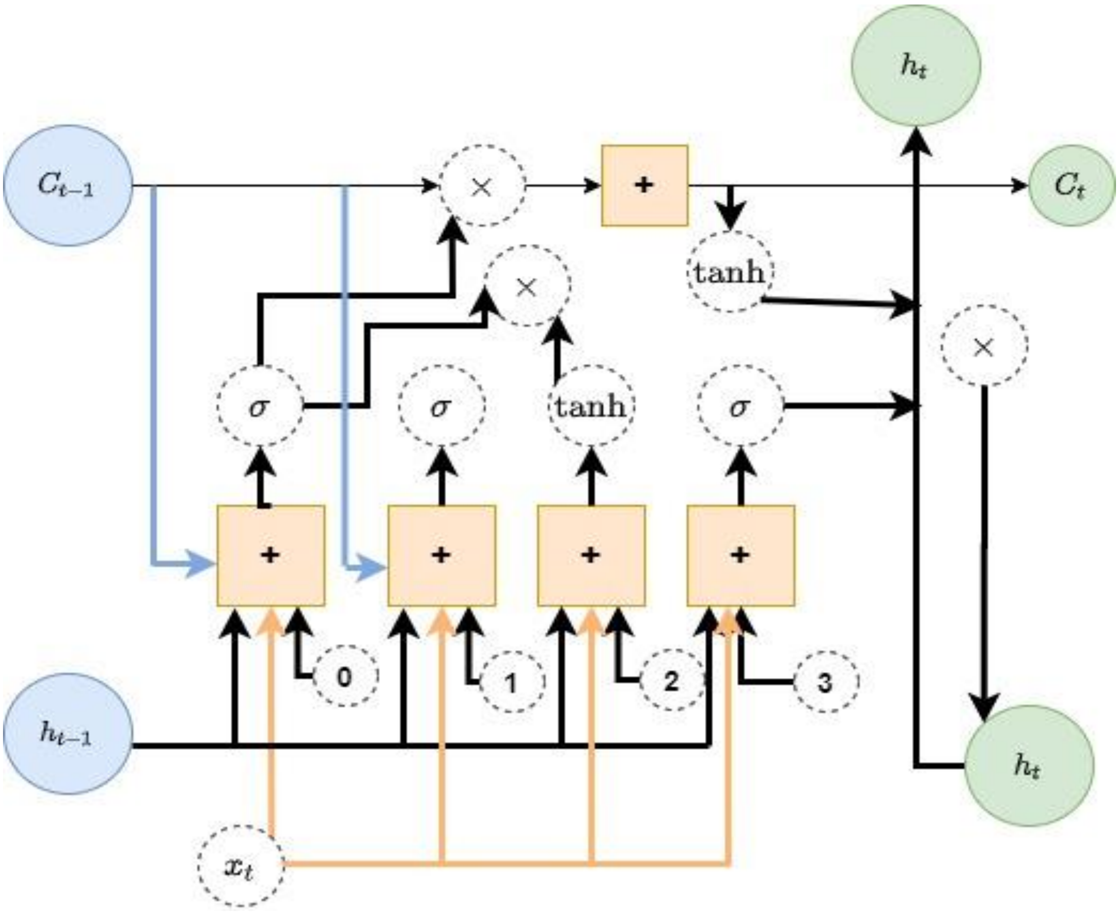



Fig. 4: LSTM basic architecture

### 4.2 Proposed Architecture Evaluation

In LSTM, the cell $c^T$ at time $T$ calculates the gradient $h^T$ and next cell state $c^{T+1}$ or the flow can be reversed in which the $c^T$depends upon the $c^{T+1}$ and $c^{T+1}$. These flows are stated as the forward pass and backward pass, respectively. In single flow LSTM, either forward or backward flow is used. LSTM can also be trained forward and backward at a time, which means it will consider the past and future values collectively to train the network. In our work, we can take this advantage and propose here to use biLSTM in which data flows in both directions at a time and learns faster. BiLSTM can help us make better use of both past and future data series at each time step. In fact, biLSTM consists of forward and backward LSTMs, which allows for a complete understanding of the characteristics of time-series data. In another word, biLSTM could avoid the blindness that propagation in unidirectional may cause. As a result, biLSTM is employed to process the input data to reflect the vibrant change in the video sequence and captures the feature more sensitively than unidirectional LSTM models.

The LSTM architectures vary on the basis of the components present at distinct layers. The number of memory units in the hidden layers, activation function, and a number of stacked layers are variables in any LSTM structure. The classification accuracy is affected by their various combinations. We test three different layers' arrangement of LSTM for various memory unit size. One layer arrangement is a conventional pattern with 32 memory units in LSTM. Another arrangement is tested with 128 number of memory units. The third proposed stacked-layer arrangement using multiple layers of biLSTM is shown in Fig. 5. For the biLSTM layers, the output size gets doubled to LSTM due to both forward and backward pass in each iteration. In the proposed scheme, we made the network deeper and complex to improve the prediction. The addition of a max-pooling layer reduces the features' dimension. To avoid overfitting, features are normalized and mapped into space of [0,1] by batch normalization layer. This set of layers with biLSTM is repeated twice in our proposal. Table 1 lists these three network architectures.

Table 1: Proposed Deep Network and Other tested LSTM architectures

| Sequence No | Layers | LSTM 1 | | | LSTM 2 | | | Proposed architecture | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Layer Presence | Input Shape | Output Size | Layer Presence | Input Shape | Output Size | Layer Presence | Input Shape | Output Size |
| 1 | LSTM/biLSTM | LSTM | 9 | 32 | LSTM | 9 | 128 | biLSTM | 9 | 256 |
| 2 | MaxPooling | ✕ | - | - | ✕ | 128 | 64 | ✔ | 256 | 128 |
| 3 | Batch normalization | ✔ | 32 | 32 | ✔ | 64 | 64 | ✔ | 128 | 128 |
| 4 | BiLSTM | ✕ | - | - | ✕ | - | - | ✔ | 128 | 128 |
| 5 | Batch normalization | ✕ | - | - | ✕ | - | - | ✔ | 128 | 128 |
| 6 | Dense | ✕ | - | - | ✕ | - | - | ✔ | 128 | 256 |
| 7 | Dropout | ✔ | 16 | 16 | ✔ | 64 | 64 | ✔ | 256 | 256 |
| 8 | Dense | ✔ | 16 | 1 | ✔ | 64 | 1 | ✔ | 256 | 1 |

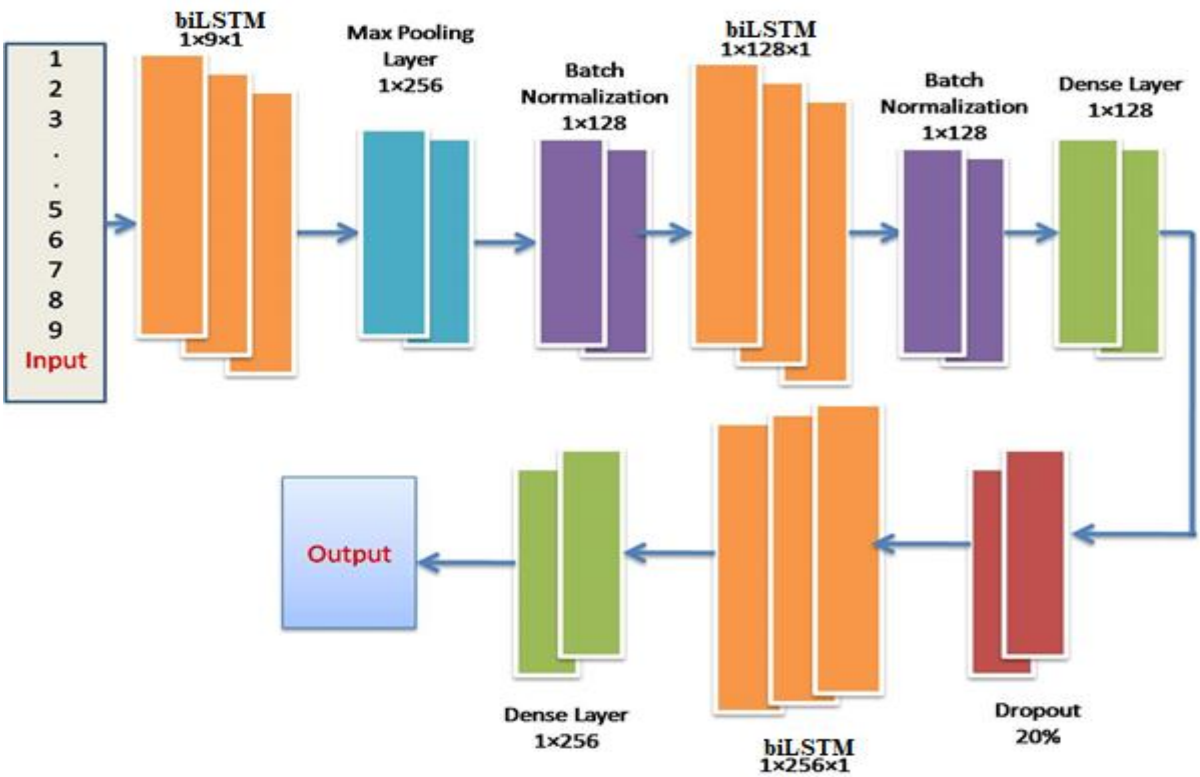
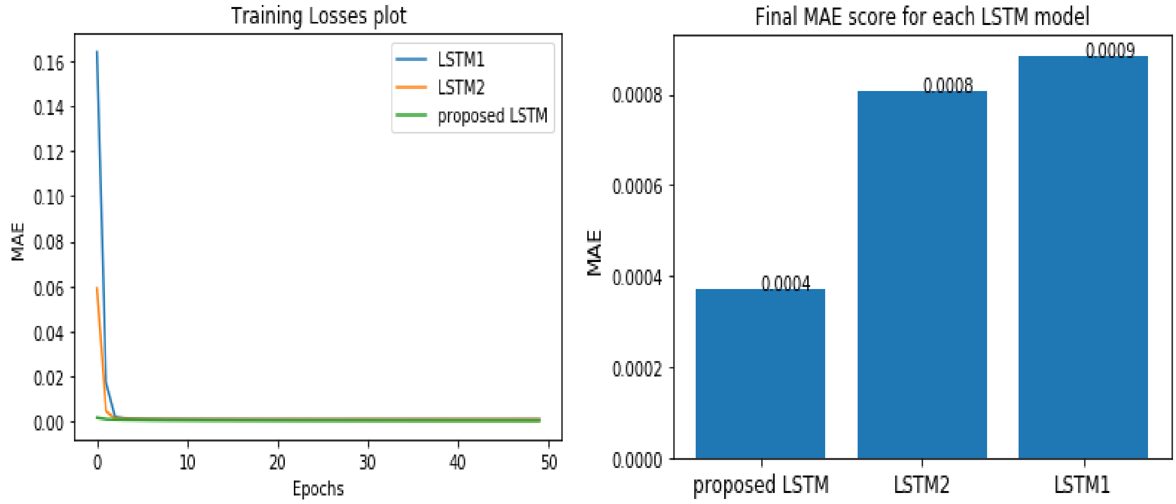



Fig. 5: Proposed stacked deep learning network for video popularity prediction

All models have trained with adam optimizer with 0.0001 learning rate, and a batch size of 128 is used. Mean Absolute Error (MAE) is the evaluation parameter for adam optimization during training. We have selected the large batch size to reduce the stochastic behavior of the model. This way, we don't need to run multiple times to consider the average performance in results. Fig. 6 (a) shows the MAE score comparison among the models with the same input data for prediction for each epoch. The proposed deep networkarrangement is reducing the MAE to 0.0004 at the last epoch. It is the least among all three models. Our network arrangement reduces the losses to 0.0018 even at the very first epoch, while others have a higher scale than it. The loss curve at each epoch for the suggested network is nearly at zero slopes. The final iterated MAE is compared in Fig. 6 (b). The input dataset is divide into 80:20 ratio for the training and testing purpose. The engagement map between watch percentage and relative engagement is built on the training set over the first 30 days. We split the dataset in time to ensure that learning is on past videos and prediction is on future videos




(a) (b)

Fig. 6: (a) Training performance comparison for all three LSTM networks on the video popularity prediction (b) MAE bar plot at last epoch for all three networks

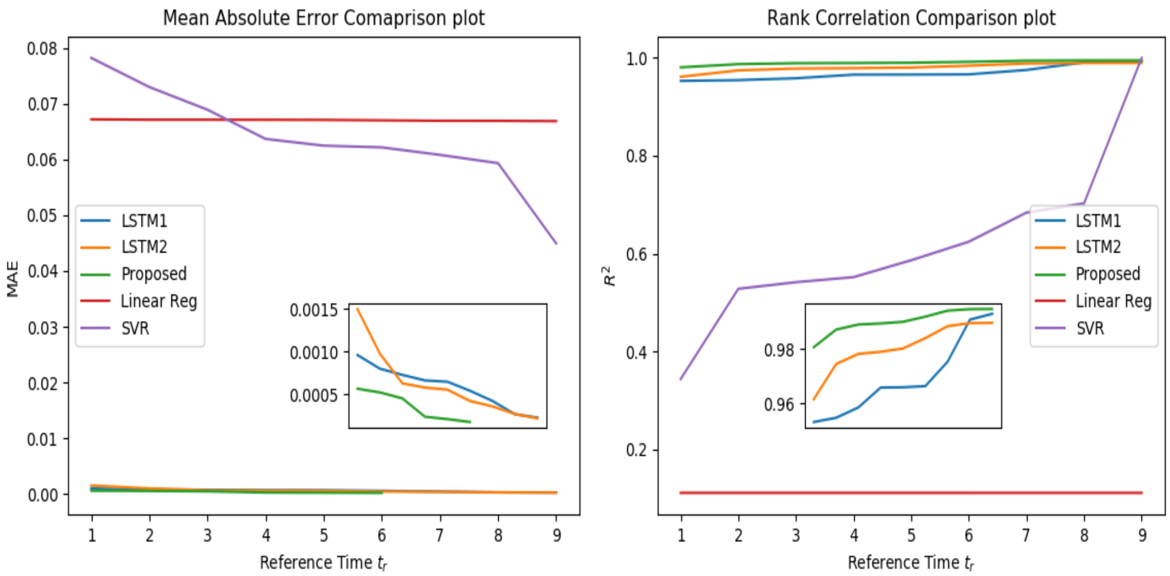
## 5.0 DISCUSSION

The time series prediction task does not follow the convention, as in other prediction applications. We can't use here cross-validation analysis as it considers that the samples are independent, but in time series, this is not the case. For cross-validation, the data is divided into testing and training sets. Evaluation of prediction is done using Mean Absolute Error (MAE) and the coefficient of determination ($R^2$) [20]. MAE and $R^2$ are the key evaluation parameters in terms of linear classification. MAE is a standard metric for average error. $R^2$ quantifies the proportion of the variance in the dependent variable that is predictable from the independent variable and is often used to compare different prediction problems. Minimum value of MAE and a higher value of $R^2$ can improve the classification accuracy and is the basic demand of accurate classification. Lower MAE and higher $R^2$ are desired. These metrics are popular due to its simplicity of application and interpretation. Wu et al. [20] proposed the linear regression with $L2$ norm. They have also tested it with K Nearest Neighbor (KNN) and Support Vector Regression (SVR) but showed down performance. LSTM has been the favorite choice of researchers for time series prediction due to its forward and backward pass flow for model training [21], [22]. The time series prediction is univariate (prediction dependent upon a single variable) and multivariate (prediction by multiple variables). Trzciński et al. [17] followed the univariate and multivariate linear regression prediction method. However, they also proposed SVR model based on the Radial Basis Function (RBF) for multivariate analysis. Though [17] worked on a different dataset, these algorithms are implemented and also validated. Since our data is multivariate, we considered the multivariate work only. Table 2 shows the performance comparison among the state of art methods with our scheme. Results shown here are for next day popularity prediction for testing data. These data samples were not used for the training of the network. It is noticeable that the proposed scheme has reduced the MAE loss upto 92% from linear regression [20], 91% from SVR [17]. This great improvement is from regression models. We also tested other deep learning architectures like conventional LSTM1 and LSTM2. The improvements from these are also recorded upto 14% and 13.6% respectively. Previously researchers also experimented with the hybrid LSTM networks, which gets the advantages of the CNN layer as the feature layers with LSTM [21], [22], [23]. We replicated their model also for testing on our data, and results are tabulated in Table 2.

Table 2: Performance comparison with the existing state of art

| | Linear Regression [20] | Support Vector Regression [17] | LSTM1 | LSTM2 | CNN-LSTM [21] | Proposed |
|---|---|---|---|---|---|---|
| MAE | 0.069 | 0.067 | 0.0004 | 0.00029 | 0.0004 | 0.00025 |
| $R^2$ | 0.11 | 0.37 | 0.9690 | 0.9703 | 0.98 | 0.9732 |

We tested the proposed scheme with previous works for the different range of reference time ($t_r$). The reference time is the time after the publishing of the video on the platform. Table 2 lists the result with $t_r = 1$. Evaluation parameters are plotted in Fig. 7 for $t_r \in [1,9]$. Deep learning methods outperform algorithms in [17], [20]. The linear regression even doesn't show any significant change with $t_r$. Both MAE and $R^2$ are following the general conviction. Since the linear regression and RBF [20] methods have a larger scale for MAE and $R^2$ with respect to proposed deep network; we have to plot the inset axis for other deep learning schemes' plots. The proposed scheme gives the lowest MAE for the same set of data, while due to more memory units, LSTM2 outperforms LSTM1. A similar performance is noted for the coefficient of determination. More is the previous metrics available for a video; higher is the $R^2$. For $t_r = 9$, the determination coefficient can achieve a value of 0.9948. Fig. 8 shows the training losses for the proposed method for various $t_r$, with an increase in $t_r$, the losses are decreasing. It validates that more are the days to video publishing; more would be the prediction accuracy.




(a) (b)

Fig. 7: Popularity Prediction results for a different time stamp after video published on YouTube. (a) MAE comparison (b) Coefficient of determination comparison

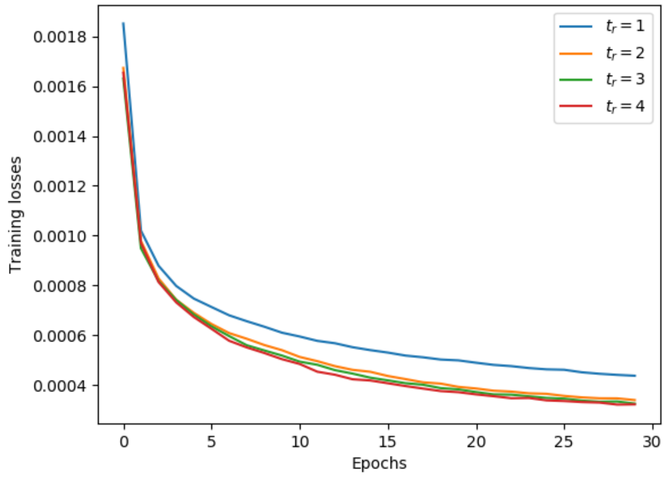



Fig. 8: Performance curve for training losses in proposed biLSTM stacked layers at $t_r \in [1,4]$

## 6.0 CONCLUSION

In this article, the author has proposed a novel stacked biLSTM network to predict the YouTube popularity on the basis of views for a database of 5331204 videos abstracted from 1257412 YouTube channels. The attributes collected from the data are temporal and video engagement features. The designed deep learning network is evaluated on the basis of mean absolute error and coefficient of determination as parameters of metric. The work is compared with the existing state of arts on the same metric. The proposed method shows an improvement of 92 % in linear regression and 14 % as compared to other LSTM architectures. The results do stand out in case of video popularity prediction, yet there is room for further improvement. The proposed scheme has focused on short-term prediction as in a long interval of time. In the future, author intend to work for more visual features for video prediction and build a hybrid or ensemble based deep learning model.

## REFERENCES

[1] H. Jeon, W. Seo, E. L. Park, S. Choi, "Hybrid Machine Learning Approach to Popularity Prediction of Newly Released Contents for Online Video Streaming Service". arXiv preprint arXiv:1901.09613, 2019.

[2] G. Chen, Q. Kong, N. Xu, W. Mao, "NPP: A neural popularity prediction model for social media content", *Neurocomputing*, Vol. 333, 2019, pp. 221-230.

[3] C. Zhu, G. Cheng, K. Wang, "Big data analytics for program popularity prediction in broadcast TV industries". IEEE Access Vol. 5, 2017, pp. 24593-24601.

[4] A. Bielski, and T. Trzcinski, "Understanding Multimodal Popularity Prediction of Social Media Videos with Self-Attention". IEEE Access, Vol. 6, 2018, pp. 74277-74287.

[5] Z. Bao, Y. Liu, Z. Zhang, H. Liu, J. Cheng, "Predicting popularity via a generative model with adaptive peeking window". *Physica A: Statistical Mechanics and its Applications*, Vol. 522, 2019, pp.54-68.

[6] C. Ma, Z. Yan, C. W. Chen, "LARM: A lifetime aware regression model for predicting YouTube video popularity" *In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ACM, 2017, pp. 467-476.

[7] L. Huang, B. Ding, A. Wang, Y. Xu, Y. Zhou, X. Li, "User Behavior Analysis and Video Popularity Prediction on a Large-Scale VoD System". *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*. Vol. 14 No. 3, 2018, pp. 1-24.

[8] P. Jing, Y. Su, L. Nie, X. Bai, J. Liu, M Wang, "Low-rank multi-view embedding learning for micro-video popularity prediction". IEEE Transactions on Knowledge and Data Engineering, Vol. 30, No. 8, 2017, pp.1519-1532.

[9] H. Dou, W. X. Zhao, Y. Zhao, D. Dong, J. R. Wen, E. Y. Chang, "Predicting the Popularity of Online Content with Knowledge-enhanced Neural Networks", *In ACM KDD*, 2018.

[10] X. Chen, J. Chen, L. Ma, J. Yao, W. Liu, J. Luo, T. Zhang, "Fine-grained video attractiveness prediction using multimodal deep learning on a large real-world dataset". *In Companion Proceedings of The Web Conference*, 2018, pp. 671-678.

[11] Q. Wu, C. Yang, H. Zhang, X. Gao, P. Weng, G. Chen, "Adversarial training model unifying feature driven and point process perspectives for event popularity prediction". *In Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp.517-526.

[12] A. Martin, S. Chintala, L. Bottou, "Wasserstein generative adversarial networks". *In International conference on machine learning*, 2017, pp. 214-223.

[13] A. Negi, A.N.J. Raj, R. Nersisson, Z. Zhuang, M. Murugappan, "RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks". *Arabian Journal for Science and Engineering,* Vol. 45, 2020, pp. 6388-6410.

[14] L. Rao, J. Yang, J., "Speech Dereverberation Based on Improved Wasserstein Generative Adversarial Networks". *Journal of Physics: Conference Series*, Vol. 1621, No. 1, 2020, pp. 1-8.

[15] C. Wu, B. Ju, Y. Wu, N.N. Xiong, S. Zhang, "WGAN-E: A Generative Adversarial Networks for Facial Feature Security". *Electronics*, Vol. 9, No. 3, 2020, pp. 1-20.

[16] J. Feng, S. Qi, "Motion Deblurring in Image Color Enhancement by WGAN". *International Journal of Optics*, Vol. 2020, 2020, pp. 1-15.

[17] T. Trzciński, P. Rokita, "Predicting popularity of online videos using support vector regression". *IEEE Transactions on Multimedia* Vol. 19, No. 11, 2017, pp. 2561-2570.

[18] S. Mekouar, N. Zrira, E. H. Bouyakhf, "Popularity Prediction of Videos in YouTube as Case Study: A Regression Analysis Study". *In Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, ACM, 2017, pp. 1-6.

[19] Q. Kong, M. A. Rizoiu, S. Wu, L. Xie, "Will This Video Go Viral: Explaining and Predicting the Popularity of YouTube Videos". In *Companion Proceedings of the Web Conference*, 2018, pp. 175-178.

[20] S. Wu, M. A. Rizoiu, L. Xie, "Beyond views: Measuring and predicting engagement in online videos". *In Twelfth International AAAI Conference on Web and Social Media*, 2018.

[21] F. Karim, S. Majumdar, H. Darabi, S. Chen, "LSTM fully convolutional networks for time series classification". IEEE Access, Vol. 6, 2017, pp. 1662-1669.

[22] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FSCNS for time series classification*". Neural Networks*, Vol. 116, 2019, pp. 237-245.

[23] M. Bisht, M. S. Irshad, N. Aggarwal, A. Anand, "Understanding Popularity Dynamics for YouTube Videos: An Interpretive Structural Modelling based Approach". *In 2019 Amity International Conference on Artificial Intelligence (AICAI)*, IEEE, 2019, pp. 588-592.

[24] M. Törhönen, M. Sjöblom, J. Hamari, "Likes and views: Investigating internet video content creators' perceptions of popularity". In *Proceedings of the 2nd International GamiFIN Conference, GamiFIN*, 2018.

[25] C. Koch, S. Werner, A.Rizk, R. Steinmetz, "MIRA: Proactive Music Video Caching using ConvNet-based Classification and Multivariate Popularity Prediction". *In IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, IEEE, 2018, pp. 109-115.

[26] F. Figueiredo, "On the prediction of popularity of trends and hits for user generated videos". *In Proceedings of the sixth ACM international conference on Web search and data mining, ACM,* 2013, pp. 741-746. ACM.

[27] H. Li, X. Ma, F. Wang, J. Liu, K. Xu, “On popularity prediction of videos shared in online social networks”.*In Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp.169-178. ACM.

[28] S. D. Roy, T. Mei, W. Zeng, S. Li, “Towards cross-domain learning for social video popularity prediction”. IEEE Transactions on multimedia, Vol. 15, No. 6, 2013, pp.1255-1267.

[29] M. C. Lee, M. Ji, A. F. Molisch, N. Sastry, “Performance of caching-based D2D video distribution with measured popularity distributions”. *arXiv preprint arXiv:1806.05380*, 2018.

[30] A. L. Jia, S. Shen, D. Li, and S. Chen, “Predicting the implicit and the explicit video popularity in a User Generated Content site with enhanced social features”. Computer Networks Vol. 140, 2018, pp. 112-125.

[31] P. Yang, N. Zhang, S. Zhang, L. Yu, J. Zhang, X. S. Shen, “Content popularity prediction towards location-aware mobile edge caching”. *IEEE Transactions on Multimedia*, Vol. 21, No. 4, 2018, pp.915-929.

[32] L. Tang, Q. Huang, A. Puntambekar, Y. Vigfusson, W. Lloyd, K. Li, “Popularity prediction of facebook videos for higher quality streaming”.*In 2017 {USENIX} Annual Technical Conference ({USENIX}{ATC}*, 2017, pp. 111-123.

[33] B. Wu, W. H. Cheng, Y. Zhang, Q. Huang, J. Li, T Mei, “Sequential prediction of social media popularity with deep temporal context networks”, *arXiv preprint arXiv:1712.0444*, 2017.

[34] J. Wu, Y. Zhou, D. M. Chiu, Z. Zhu, “Modeling dynamics of online video popularity”, IEEE Transactions on Multimedia, Vol. 18, No. 9, 2016, pp. 1882-1895.

[35] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, W. V. Dolen, “Multimodal popularity prediction of brand-related social media posts”.*In Proceedings of the 24th ACM international conference on Multimedia*, ACM, 2016, pp. 197-201.

[36] C. Li, J. Liu, and S. Ouyang, “Characterizing and predicting the popularity of online videos”. IEEE Access, Vol. 4, 2016, pp. 1630-1641.

[37] Y. Wu, Z. Meng, S. Palaiahnakote, T. Lu, “Compressive Sensing Based Convolutional Neural Network for Object Detection”. *Malaysian Journal of Computer Science*, Vol. 33, No. 1, 2020.

[38] N. Khairudin, N. M. Sharef, S.A.M. Noah, N. Mustapha, “Embedded learning for leveraging multi-aspect in rating prediction of personalized recommendation”. *Malaysian Journal of Computer Science*, 2018, pp. 31-47.

[39] E. B. Abdelkrim, M. A. Salahuddin, H. Elbiaze, R. Glitho, “A hybrid regression model for video popularity-based cache replacement in content delivery networks”. *In 2016 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2016, pp. 1-7.

[40] Z. Tan, Y. Wang, Y. Zhang, J. Zhou, “A novel time series approach for predicting the long-term popularity of online videos”. IEEE Transactions on Broadcasting, Vol. 62 No. 2, 2016, pp.436-445.