International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

Optimized Ensemble Approach for Multi-model Event Detection in Big data

K. Swapnika[1*], D. Vasumathi[2]
[*1]Research Scholar, Computer Science and Engineering,
Jawaharlal Nehru Technological University, Hyderabad, Telangana 500085, India.
*Email: swapnika.griet@gmail.com
[2]Professor and Head of the Department, Computer Science and Engineering,
Jawaharlal Nehru Technological University, Hyderabad, Telangana 500085, India.
Email: vasukumar_devara@jntuh.ac.in

Abstract: Event detection acts an important role among modern society and it is a popular computer process that permits to detect the events automatically. Big data is more useful for the event detection due to large size of data. Multimodal event detection is utilized for the detection of events using heterogeneous types of data. This work aims to perform for classification of diverse events using Optimized Ensemble learning approach. The Multi-modal event data including text, image and audio are sent to the user devices from cloud or server where three models are generated for processing audio, text and image. At first, the text, image and audio data is processed separately. The process of creating a text model includes pre-processing using Imputation of missing values and data normalization. Then the textual feature extraction using integrated N-gram approach. The Generation of text model using Convolutional two directional LSTM (2DCon_LSTM). The steps involved in image model generation are pre-processing using Min-Max Gaussian filtering (MMGF). Image feature extraction using VGG-16 network model and generation of image model using Tweaked auto encoder (TAE) model. The steps involved in audio model generation are pre-processing using Discrete wavelet transform (DWT). Then the audio feature extraction using Hilbert Huang transform (HHT) and Generation of audio model using Attention based convolutional capsule network (Attn_CCNet). The features obtained by the generated models of text, image and audio are fused together by feature ensemble approach. From the fused feature vector, the optimal features are trained through improved battle royal optimization (IBRO) algorithm. A deep learning model called Convolutional duo Gated recurrent unit with auto encoder (C-Duo GRU_AE) is used as a classifier. Finally, different types of events are classified where the global model are then sent to the user devices with high security and offers better decision making process. The proposed methodology achieves better performances are Accuracy (99.93%), F1-score (99.91%), precision (99.93%), Recall (99.93%), processing time (17seconds) and training time (0.05seconds). Performance analysis exceeds several comparable methodologies in precision, recall, accuracy, F1 score, training time, and processing time. This designates that the proposed methodology achieves improved performance than the compared schemes. In addition, the proposed scheme detects the multi-modal events accurately.

Keywords: Event detection, pre-processing, Feature extraction, Feature fusion, Feature selection, Multi modal generation, Classification.

I. INTRODUCTION

Event detection is a computer process that allows the automatic identification of significant activities, through the analysis of social media data [1]. A notable incident that occurs at a certain location and time is referred to as an event. Because it allows for instant access to information about current events and public opinion. Event detection in social media streams is an important research subject. The numerous overwhelming news through a news organization or agency are long gone in which someone could destroy information that they did not want other people to be aware of. Now there are numerous social media platforms, this is no longer feasible. To obtain significant insight into hot topics and events on social media [2], one Big Data task that has evolved is event detection. More specifically, social media has a significant influence on the development of big data, and it is a useful tool for comprehending the enormous amount of data produced on microblogs [3]. The ability to perform iterative data discovery, which results in analysis, detection, and information extraction such as popular events, is what big data most importantly permits. According to this challenge, in particularly concentrate on this work on the topic of automatic online event detection [4] on Twitter microblogs by fusing a big data analytics environment [5] with Twitter analytics to produce a novel approach that can improve event detection within the big data space. Big data demonstrated that social media data is useful for identifying data dissemination features and spotting earthquake myths. This motivates us to research the issue of event detection [6], which in these circumstances is a fascinating and crucial endeavor. In reality, because of tweets unique properties, event detection methods [7] intended for documents cannot be rigorously implemented.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

Rapid data processing is required to address emergency situations [8] including natural disasters and human-caused casualties, which cause spikes in social media network information rates [9]. Big data processing solutions are required because social media networks produce enormous amounts of data during emergencies that typical data processing methods cannot handle. Massive volumes of data may be researched, stored, and monitored using a variety of increasingly complex techniques using big data. It enables decision-makers to study and comprehend a set of information to make the best decisions in emergencies. To mitigate the disastrous effects of such incidents, a comprehensive emergency event detection system is required. Big data and social links [10] have a substantial effect on how quickly emergency occurrences are discovered. The management, observation, analysis, and detection of emergency incidents may benefit from the use of social network data. Nevertheless, the location of emergency situations must be precisely identified while gathering real-time data [11] from social networks.

Event detection is crucial for learning data semantic [12] methods for data summarizing, retrieval, and indexing. In order to identify when data analysis occurs, a lot of study has been done in this area. Many prior event detection techniques train their event detection models using labelled samples, domain knowledge characteristics and videos. The deployment of event recognition systems is hindered by different factors such as backdrop clutter [13], ambiguous visual cues, the semantic gap among decreased level of features and high level events of different classes of films, ambiguous visual signals, and changing alternations of camera movements. Big data seems to be more useful for event detection due to the enormous quantity of storing capacity. In addition, prior methods of event detection have a strong domain bias. To find events in massive amounts of data sources including text, images, video and audio recordings, multimodal event detection [14] techniques are currently developed. Machine learning techniques have a difficult time detecting events in a particular domain. Many processing layers are used in deep learning models to understand data format at various abstraction levels.

Due to the requirement to sense the city at the micro-level, make wise decisions, and take appropriate actions, all under strict time constraints, big data technologies would play a crucial role in supporting smart city systems and applications. Social media has completely changed the communities, and as it gathers data on people and their spatial and temporal experiences in and around their homes. It is progressively emerging as a vital component of smart societies. Certain methods for finding spatiotemporal occurrences in event detection, such as cyclones and earthquakes, are available. In this discipline, methods regard people who text about occurrences on social media as sensors. Events are also discovered using space-time scan statistics (STSS) out without aid of the text utilizing only space and time [15]. STSS perceives text in a space-time cube, which moves a cylindrical window over all imaginable space-time locations with a height (time) and variable radius (space). The ST-DBSCAN method [16] groups texts across space and time dimensions using geo-tags and timing. By mapping term frequency-inverse document frequency (TF-IDF) [17] feature dependent words into spatiotemporal space, multiple researchers classified texts into different events using machine learning approaches [18]. The ST-DBSCAN algorithm groups tweets based on timing and geo tags in both time and location. IDS (intrusion detection system) uses data from a different sources to accomplish the validation. Some of them gather data from network traffic statistics [19], while others look for particular events that are either predefined or discovered by ML methods. For event detection, current deep learning (DL) frameworks like bidirectional long short term memory (BiLSTM), convolutional neural network (CNN), Deep belief network (DBN), capsule auto encoder (CAE), ResNet and so on performs better than existing machine learning (ML) techniques like ST-DBSCAN, support vector machine, decision tree, logistic regression, k-nearest neighbor classifier and so on [20]. In existing, various ML approaches are used for better event detection for various big data applications. However, it is challenging to predict the events when the number of processing data is increased. To enhance the accuracy and robustness of event detection, a further advanced methodology is needed. Instead of using other existing learning based approaches, ensemble learning framework can detect the events accurately. Henceforth, an optimized ensemble approach is used for the detection of events accurately.

Motivation: Event detection in big data acts a significant part in various applications. Currently, number of researchers are used different approaches to detect the data events. Moreover, only a few amount of works are done in big data event detection. The data analysis based approaches are crucial meanwhile they permits for the scalability of network in regards of information analytics. The multi-modal event detection is complex process and numerous approaches are used for diversity of use-cases. Most of the current schemes are focused only on unimodal event identification. Therefore, an effective methodology is in need for an accurate detection of multi-modal events. The existing approaches are good in event classification, though the performance is enhanced further with an improved deep learning based schemes. The ensemble deep learning framework is advanced and it can solve the issues present in the existing approaches in terms of effectiveness. Therefore, the proposed event detection used an optimized ensemble deep learning framework for an accurate identification of various events. The proposed framework can enhance the effectiveness

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

approach by increasing the efficiency and accuracy of framework. Subsequent are the main contributions of the suggested approach:

- To attain an efficient multi-modal event identification process with an optimized ensemble learning technique.
- Improve the overall performance while implementing the event detection scheme for text, image, and audio information.
- An effective methods for text, image, and audio data pre-processing is utilized to obtain precise outcomes for the big data event detection.
- To evaluate the event detection system's performance in comparison to its existing performance using statistical measures like f1-score, recall, precision, training time, accuracy, and testing time.

The suggested methodology is summarized as: Section 2 contains a comprehensive description of related works; Section 3 provides an explanation of the proposed technique; Sections 4 provide an explanation of the results and discussion and the proposed methodology is concluded in section 5.

II. RELATED WORK

Distributed machine learning over Apache Spark was used to implement the method proposed by Ebtesam Alomari et al. [21] to automatically identify traffic occurrences from texts written in the Arabic language. The programme is termed Iktishaf+ (an Arabic word for discovery). There were nine parts to the tool, which used a number of different technologies such Apache Spark, Parquet, and Mongo DB. For the Arabic language, Iktishaf+ uses a lightweight stemmer they developed. In this work, we also developed a location extractor that they utilize to extract and visualize spatiotemporal data regarding observed events. Support vector machines, logistic regression, and naive Bayes -based classifiers can be used to recognize and assess a variety of genuine events that have occurred like a fire in Jeddah, an accident in Riyadh, and rain in Makkah. The results depicts how social media may effectively identify significant events when there is no prior knowledge about them.

In order to respond quickly when such catastrophic circumstances arise, Khalid Alfalqi and Martine Bellaiche [22] proposed a method where an efficient emergency event detection ensemble model (EEDEM) was necessary. In order to offer a fresh method of learning the actual location of an emergency occurrence, it also combines Snap chat maps. Social network data, like that from snap and Twitter chat, allow to analyze, monitor, detect and manage emergency occurrences. Combining big data and social media also assist to speed up the emergency event detection system. The main objective was to provide an innovative and effective big data oriented EEDEM to locate precise position of events using information gathered from social networks like Snap chat and Twitter while combining machine learning (ML) and big data (BD). Additionally, established study evaluates the effectiveness of the proposed ensemble detection method and five ML base models. With a very high accuracy of 99.87%, the proposed ensemble technique surpassed the other basic models. Moreover, with an acceptable training time, the suggested models get a high accuracy of 99.70% for the decision tree and 99.72% for LSTM correspondingly.

Ebtesam Alomari *et al.* [23] introduced Iktishaf for detecting traffic based events from Twitter information in Saudi Arabia. It constructs numerous classifiers with three machine learning (ML) techniques to detect eight various types of events. The classifiers were tested beside external sources and frequently accepted criteria. Text preparation, event identification, and feature space were enhanced. Without prior knowledge, they are able to identify events like the fire in Riyadh, KSA National Day, Taif, the opening of the Al-Haramain railway and rain in Makkah by using 2.5 million tweets. They were not aware of any research that analyses Arabic tweets for traffic incidents using big data methods. Iktishaf offers hybrid human-ML approaches and was a great example of combining massive data processing, human cognition and AI theory, to a real-world issue.

Yasmeen George *et al.* [24] offer a social media-based online Spatio temporal event identification framework that can identify events at various spatial and time solutions. To begin, a quad-tree method was used to divide the geographical space into multiple scale zones depends on the amount of social media information to handle the difficulty of uncertain spatial resolution of events. Then, a statistical unsupervised approach involving Poisson distribution and a smoothing method was used to detect places with unusual amount of social posts. Furthermore, the duration of an event was precisely approximated by combining occurrences that occur in the same place at sequential periods. To remove fake, inaccurate events, and spam a post-processing layer was implemented. In order to evaluate the authenticity and accuracy of detected occurrences, they also incorporate basic semantics by using social media platforms. Different social media datasets, including those from Flickr and Twitter, were used to test the suggested technique in London, Paris, New York and Melbourne among other locations. They contrast the outcomes with two benchmark algorithms that use the clustering method and a fixed geographical space split to verify the efficacy of the proposed strategy. They manually compute recall and precision to evaluate performance. They also suggest a new quality metric called the strength index, which determines the accuracy of the reported event automatically.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
Article Received: 24 March 2023 Revised: 20 May 2023 Accepted: 02 June 2023

To build an intrusion detection platform and gather massive amounts of data for intrusion detection, Hye-Min Lee and Sang-Joon Lee [25] recommended using DL and CNN (Convolutional Neural Networks). By gathering and examining user visit logs and linking to big data, they develop an intelligent big data platform that gathers data. For intrusion detection, they want to gather a sizable amount of data and develop a system depending on the CNN architecture. The performance of the intrusion framework was assessed in developed work utilizing the KDD99 database created the DARPA in 1998, and real attack classes were tested using KDD99's DoS, R2L and U2R utilizing four probing approaches.

Khalid Alfalqi et al. [26] developed an ensemble framework for the emergency event detection in big data. The framework combines the snapshots to present an innovative approach to predict the emergency events accurately. Different events were gathered from the social networks like Snapchat and Twitter. The integration of big data for the event detection was performed using an innovative machine learning approaches. The developed framework attains good performance in event detection. The presented scheme attains good performance in event identification. Still, the performance of event detection can be enhanced further utilizing an improved form of schemes.

Alan D. Smith et al. [27] presented a framework for the event detection in educational big data records. The education oriented event detection considers the R- tree based, query based big data analysis model. The developed model can predict the minor variations in the simulated data. Here, data clustering approach was applied to cluster the category of educational data. The Mahalanobis distance was estimated to validate the distance between events to categorize the events accurately. The performance needs to be enhanced further by utilizing an enhanced schemes.

Problem Formulation: There is a bare minimum of work is done in order to identify the events. The methods used to find events must be precise. To improve decisions, the data must be accurately retrieved. However, most of current networks offer less precision for event detection processes. The challenges in processing time and complexity of the existing approaches is another restriction. Since they enable systems to scale in terms of information analytics and management, big data approaches are essential. Big data is used in a diversity of works to identify events. This work aims to perform for classification of diverse events using optimized ensemble learning model.

III PROPOSED METHODOLOGY

This paper presented an optimization-based ensemble learning for event detection in big data. Three models are built for processing text, image, and audio before the multi-modal event data, which includes text, image, and audio, is provided to the user devices from the cloud or server. In the text model generation, pre-processing is done by Missing value imputation and data normalization. For the feature extraction, integrated N-gram approach and model generation is done by Convolutional two directional LSTM (2DCon_LSTM). In the image model generation, pre-processing is done by Min-Max Gaussian filtering (MMGF). Feature extraction is carried out using the VGG-16 network model and model generation is done by Tweaked auto encoder (TAE) model. The processing step involved in the audio model generation are pre-processing, feature extraction and model generation by Hilbert Huang transform (HHT), Discrete wavelet transform (DWT), Attention based convolutional capsule network (Attn_CCNet) respectively. The work flow of the proposed approach is shown in figure 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

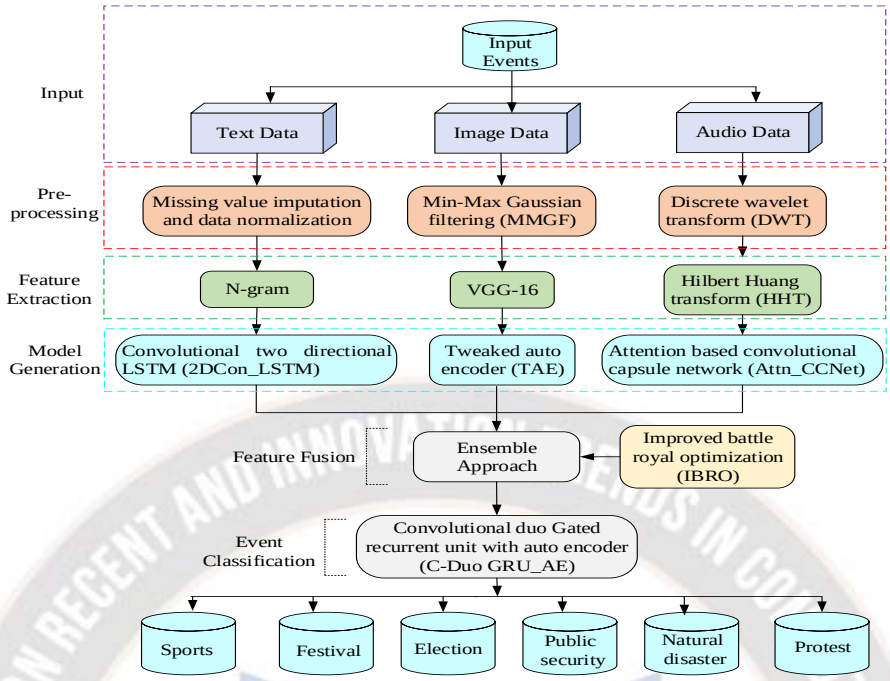


Figure 1: Work flow of the proposed methodology

The feature ensemble approach is used to combine the features that were gathered from the text, image, and audio models. The improved battle royal optimization (IBRO) technique is used to train the best features from the fused feature vector. A deep learning model called Convolutional duo Gated recurrent unit with auto encoder (C-Duo GRU_AE) is used as a classifier. Finally, various events are categorised, and the global model is then provided to user devices with high security and superior decision-making capabilities.

A. Text model generation

1. Pre-processing

The procedure of pre-processing by cleaning the text for classification is performed initially. The text which is obtained from internet typically have a lot of useless content and noise like ads, scripts, and HTML tags. During the training process, knowledge discovery becomes particularly challenging if there is unsuitable information present or noisy and unreliable data. Despite the fact that pre-processing tasks like data transformation and cleaning can take a long time to complete, the data become more trustworthy and solid conclusions have been acquired. Consequently, pre-processing should be done before extraction. Here, Missing value imputation and data normalization is used for pre-processing of text data.

I) Missing value imputation

When data mining techniques are used on a particular data collection, missing values are one of the key components. Missing values in a data set could occur for a variety of causes, including human mistake, hardware issues, etc. It should be emphasised that before analysis, the missing values should be carefully dealt. Consequently, the data extracted from a data base with missing values may point decision-makers in the wrong direction.

II) Data normalization

2. Feature Extraction

III) N- Gram approach

The feature extraction (FE) method focused to eliminate the irrelevant data from text. N-Gram represents the method utilized here for feature extraction. N-gram features are commonly used in tasks involving text content classification. According to table 1, the n-gram characteristics can be separated into word and letter n-gram features. Unigrams, bigrams, trigrams, etc., are terms used to describe n-grams of length 1 (also known as unigrams), n-grams of length 2 (also known as bigrams), and so on.

TABLE 1: Example for N-gram

Sentence/ N-gram	This is Big Data Book
Unigram	This, is, Big, Data, Book
Bigram	This is, is Big, Big Data, Data Book
Trigram	This is Big, is Big Data, Big Data Book

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

Typically, people use N-gram parsers to scan articles, break them up into phrases, and count the grammes in combination with the outputs of the compression unit.

C. Process for Generation of text model

IV) Convolutional two directional LSTM (2DCon_LSTM)

This section covers the input and output, convolutional phrase encoder, and recurrent document decoder of LSTM-CNN framework. The initial four layers of the text classifier depends on LSTM and CNN are the LSTM or one of its derivatives layer, the input layer, convolutional network layer, and the softmax layer. Generation of text model using 2DCon_LSTM is depicted in figure 2.

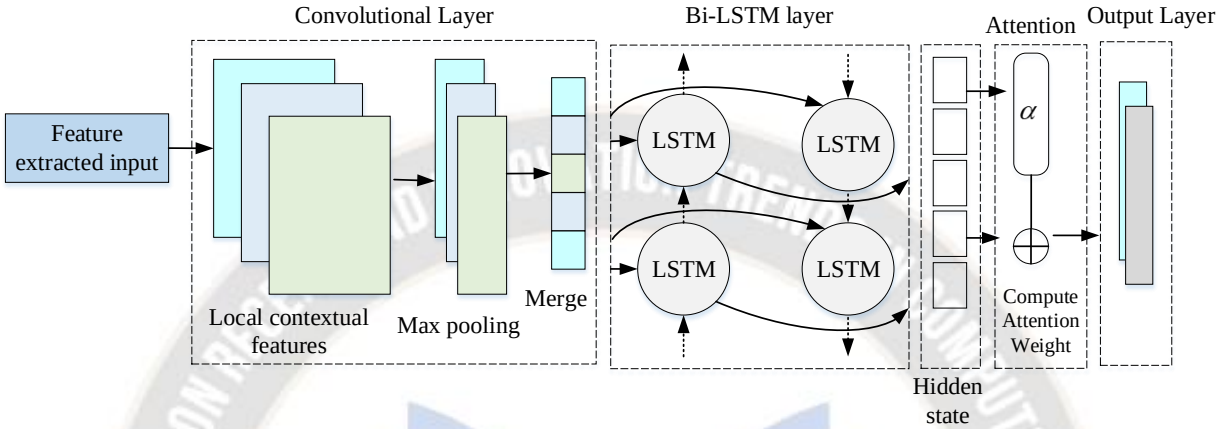


Figure 2: Generation of text model

The input feature is given as input to the convolutional layer for the generation of text. The kernel function, also referred to as the filter, is the central component of convolutional operation. The kernel function in text generation processing ensures the reliability of word as finest granularity in text. Its width is normally equivalent to the width of unique matrix and it only slides in the top and lower dimensions. There are two types of padding techniques in the kernel function's sliding process: legal padding and zero padding, which are determined by whether the original matrix is increased by zero. The LSTM layer receives the output of the convolutional layer. As the pooling procedure will sever the sequential relationship that the LSTM needs as input, it is best to ignore this process.

The input gate, output gate and forget gate constitute an LSTM. After the forget gate has selected which data to remove from a cell state, the input gate chooses which data to be updated. The cell state could be updated once these two points have been established. The output gate, in the end, determines the network's ultimate output. The output of the LSTM layer is updated by equation (1).

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \# \qquad (1)$$

Where C_t stands for the layer's cell state. The forget gate's output is indicated by f_t. The intermediate temporary state is denoted by $\tilde{C}_t$. The previous layer's cell state is denoted by C_{t-1}, and $1 - f_t$ controls the input gates' output.

For classification, the characteristics are delivered in a completely linked manner to the softmax classifier. Specifically, Softmax is a type of function. When making a prediction, It is capable of selecting the class with highest value of probability and translating neural output to the range (0,1). The computation of softmax value is represented in condition (2).

$$P_i = \frac{e^i}{\sum je^j} \# \qquad (2)$$

Where P_i stands for the ith category's probability, e^i for the ith category's output's corresponding value and j stands for the sum of total categories.

D. Image model generation

3. Pre-processing using Min-Max Gaussian filtering (MMGF)

To get better classification results, image data must typically be pre-processed because it often contains undesired information and noise. There are many techniques used for pre-processing image data. Here, it is achieved with the intervention of Min-Max normalization and Gaussian filtering technique. In the data pre-processing procedure known as data normalisation, the scales of the characteristics are modified to have a uniform scale of measurement. By converting the value of each feature in the range of zero and one, Min-Max Normalization, also signified as (0-1) Normalization, varies Y to Y'. If the data were negative, the possible values would have ranged from -1 to 1. Equation 3 describes the Min-Max Normalization formula:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

$$Y' = \frac{Y - \min(Y)}{\max(Y) - \min(Y)} \qquad (3)$$

Where, Y' represents the normalized value, Y represents the original value and maximum value of Y and minimum value of Y denoted as $\max(Y)$ and $\min(Y)$ respectively. Since normalisation speeds up the training process, it may increase the neural networks' (NNW) classification accuracy. The normalization output is given to the Gaussian filtering technique to reduce the signal distortion. Such filters' impulse response is a Gaussian function; filters with a Gaussian function have a Gaussian impulse response. With the least amount of group delay is possible with this method. In image processing, the average value of nearby pixels is calculated using the Gaussian function using the Gaussian smoothing. The effect of noise and other illuminations is eliminated by this operator. The high frequency components in the image are removed by its function as a Gaussian low pass filter.

$$I_s(a,b) = I_G(a,b) * G(a,b) \qquad (4)$$

Where, $I_s(a,b)$ is the Gaussian noise. The green channel component is denoted as $I_G(a,b)$, Gaussian function is represented as $G(a,b)$ and the convolution is denoted as *. It can filter images by refining in light of the fact that this filter has a kernel centre so the Gaussian filtering technique was used here. The removal of noise that is typically distributed is quite effective with this filter.

4. Feature extraction using VGG-16 network model

In a character recognition system, feature extraction is carried out after the pre-processing stage. An input pattern must be accurately assigned to one of the potential output classes in order for pattern recognition to be effective. Any pattern classification process must begin with feature extraction, which tries to collect the pertinent data that defines each class. A feature vector serves as the identity of each character throughout the extraction process of features. The major objective of feature extraction is to develop feature sets that are similar for various instances of the same symbol and maximise recognition rate with the fewest number of components. Here the pre-processed image is given as the input to the VGG-16.

The Visual Geometric Group is known as VGG. Since AlexNet's design was comparable to VGG Net's, a large number of features were also present. There are 138 million parameters in this network. The archetecture of VGG-16 is depicted in figure 3.

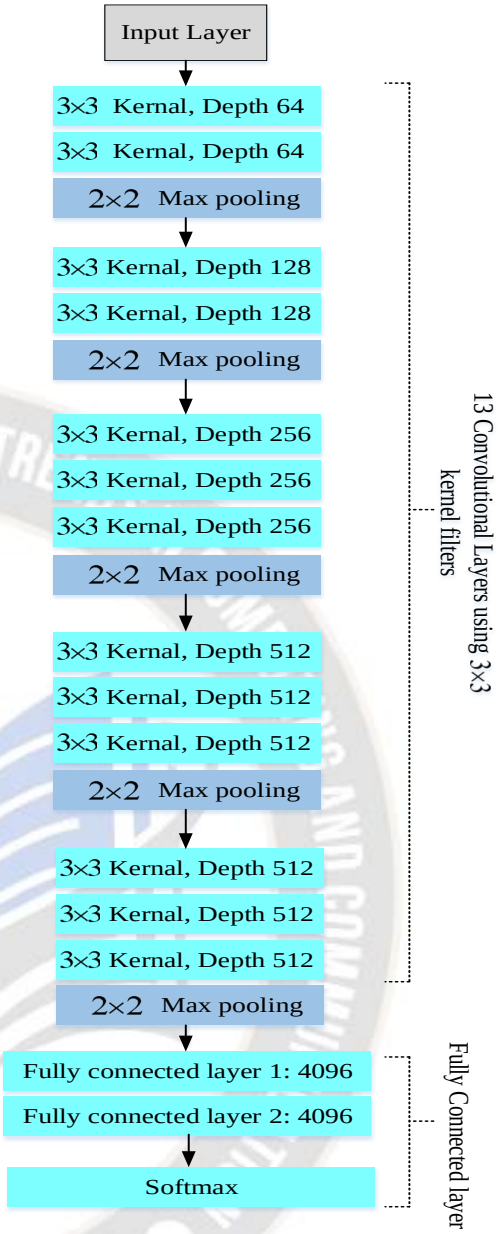


Figure 3: Archetecture of VGG-16

A 64 feature kernel with 3×3 filter size creates a first and second convolutional layers. The primary and secondary convolutional layers are used for the processing of input RGB image with a depth of 3, resulting in a 224x224x64 size change. A 3*3 filter size is used for 124 feature kernel filters that make up the next 3rd and 4th convolutional layers. The outcome is decreased to 56x56x128 as a result of these two layers being next to the max pooling layer with stride 2. The 5th, 6th, and 7th are convolutional layers with a kernel size of 3*3. 256 feature maps are used by all three. A max pooling layer with stride 2 is added after these layers. There are two sets of convolutional layers with kernel sizes of 8 and 13, respectively. In each of these collections of convolutional layers, there are 512 kernel filters. Subsequent of these layers is a max pooling with stride

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

of 1. Next, the 14th and 15th levels are fully connected and hidden layers with a combined total of 4096 units, immediately following a softmax output (sixteenth layer) of 1000 units. The output obtained from VGG-16 given to the TAE to generate the image model.

5. Generation of image model using Tweaked auto encoder (TAE)

The tweaked auto encoder (TAE) neural network is comprised of several sparse auto-encoders associated end to end. The output of modified self-encoder of preceding layer is used as an input to the following layer to generate highest level of feature demonstration from the input. The structure of tweaked auto encoder is given in figure 4.

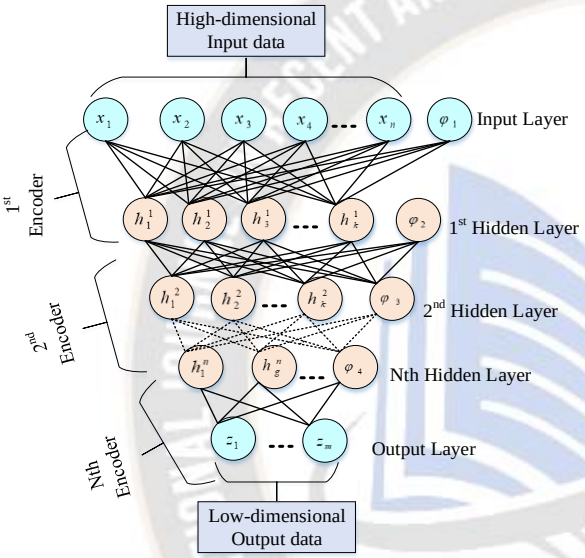


Figure 4: Architecture of Tweaked Auto encoder

The structure of TAE framework includes input, hidden and the output layer. TAE based image generation process comprised of encoding and decoding. Here, encoder is used for mapping the input information into their hidden demonstration and decoder is used for reconstructing the input information from hidden denotion. The encoding process is represented as,

$$h_k = F\left(w_1 y_k + b_1\right) \quad (5)$$

Here, F signifies the function of encoding, w_1 represents the encoder's weight matrix, b_1 represents the bias value. Consequently, the decoding process is represented as,

$$y_k = G\left(w_2 y_k + b_2\right) \quad (6)$$

Furthermore, the TAE parameter optimization to reduce the reconstruction error is described as,

$$\phi(k) = \arg\min_{\theta,\theta'} 1/k \sum_{j=1}^{k} M\left(y^j, \hat{y}^j\right) \quad (7)$$

Here, k denotes the loss function $M\left(y^j, \hat{y}^j\right) = \left\| y^j - \hat{y}^j \right\|^2$. According to the representation of figure 4, the TAE structure comprised of stacking of number of auto encoder layers. The learning is performed in each TAE layers are by the layer wise learning and all the layers are fine-tuned. At first, the TAE framework is trained by the input data and the model is generated by the learning process. The data in the prior layer is given as an input to next layers and the process of learning is repetitive until a training process completed. After the completion of hidden layers training, loss is minimized and the weights are updated for fine tuning the data. The best connection among bias and weight values for a total stacked sparse auto-encoded network is obtained by sequential training in each layer of TAE by using greedy layer-wise pre-training approach. Once the outcome of error function among both input and output information fulfils the necessary needs, the error back propagation scheme is utilized to fine tuning of TAE to produce the optimal parameter model. Here, the image gestures are generated from the extracted features by using the TAE framework. The image generation is used for the representation of framework for further processing.

E. Audio model generation

6. Pre-processing using discrete wavelet transform (DWT)

To create a reliable and suitable audio signal representation, pre-processing of the input audio signals is essential. Typically, background noise and forefront acoustic objects are present in an audio signal captured with a microphone in reality. The cause is that redundancy in signals must first be eliminated.

Here the pre-processing is done by DWT method. Noises of various kinds can distort any digital signal. It is required to remove noise from signal features before isolating them.

An expression for a signal that has been corrupted by noise is:

$$w(t) = n(t) + s(t) \quad (8)$$

Here, $w(t)$ is denoted as audio signal, $n(t)$ represents the noise contained in the audio and $s(t)$ is audio signal without noise distortion. A typical technique for removing noise from a signal is the wavelet transform. A DWT belonging to the symlet family was employed to remove the noise from the audio signals. The detail factor is determined empirically for each scenario. The wavelet transform removes noise from the audio signal in three steps. Using the DWT of a noisy signal, acquire noisy wavelet coefficients as the first step. Making a thresholding decision is the second stage. An inverse wavelet transform is used in the third stage to produce a signal that has been tidied up. The audio signal's DWT is:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

$$D_{wt}(x,y)=\frac{1}{\sqrt{2}}\sum_{j=0}^{n}w_j\int_{j}^{j+1}\varphi(\frac{T-b}{a})dt \qquad (9)$$

Here, $D_{wt}(x,y)$ represents the extracted DWT feature, audio signal distorted by the noise is denoted as w, the samples total count on the audio signal is denoted as n, the symlet is denoted as φ and the variables x, y are denoted as $x=1...n$ and $y=1...n-1$ respectively.

By selecting the threshold function and thresholding the cleaning of noise from audio signal is done in the next step. The noise in the signal is identified by threshoding value. The wavelet coefficient value at a given time is regarded as the signal if it exceeds the threshold value; else, it is regarded as the noise.

The following is the reverse DWT for getting a cleaned audio signal.

$$s(t)=\sum_{K=-\infty}^{\infty}\tilde{g}_k Y_{lowK}(t)+\sum_{m=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\tilde{h}_{mK}Y_{high_k}(t) \qquad (10)$$

Here, $s(t)$ denotes the audio signal without noise distortion, following threshold function processing, coefficients are approximated and detailed by $\tilde{g}_k$ and $\tilde{h}_{mK}$.

7. Audio feature extraction using Hilbert Huang transform (HHT)

The technique of identifying a signal's dominant and differentiating characteristics is known as feature extraction. In a much more compact manner, an appropriate feature duplicates the characteristics of a signal. Here HHT technique is used for the feature extraction of audio signal.

HHT explains the notion of an intrinsic modal functions (IMF) whilst completely maintaining all types of data in the signal by breaking a signal down into a continuous of intrinsic modal functions (IMFs). There are two parts for HHT.

By analysing the local time aspects of signal using empirical mode decomposition (EMD), a constrained number of IMFs can be found in the first section. Two requirements must be met by IMF. There are two instances when this is true: the average envelope created through the local extremes is zero, and there are either exactly as many zero crossings as there are extreme locations. A unique screening procedure is used for EMD. Calculating the sample $X(t)$ of lower and upper envelopes and determining the mean value M should come first. A new sequence is obtained by,

$$Y_1(t)=X(t)-M_1(t) \qquad (11)$$

However, $Y_1(t)$ is unable to fulfil the requirements to acquire the IMF. Consequently, numerous screenings are required. The current sequence is based on the outcome of the preceding calculation. If $Y_1K(t)$ could meet the IMF properties after the $K+1$ sifting,

$$Y_{1K}(t)=Y_{1(K-1)}(t)-M_1(t) \qquad (12)$$

The sample be able to be represented as follows when the decomposition cycle is complete:

$$X(t)=\sum_{i}^{n}C_i(t)-R_n(t) \qquad (13)$$

Here, $C_1(t)$, $C_2(t)$,...$C_n(t)$ are IMFs and residual function is $R_n(t)$. It depicts the sample's average trend.

The Hilbert transform is the next step in the handling of IMFs. Using the integral formula, the following is the result for the new data series $H_i(t)$.

$$H_i(t)=\frac{P_v}{\pi}\int_{-\infty}^{\infty}\frac{C_i(\tau)}{t-\tau}d\tau \qquad (14)$$

P_v stands for the value of the Cauchy principle there.

Initially, to extract the intrinsic mode functions (IMFs), execute EMD decomposition on each sample. So it is possible to calculate the component total and maximum amplitude of IMF1, IMF2,…., IMF5 to obtain characteristics. Next, the IMF1, IMF2, and IMF can be transformed into Hilbert space to determine the loss amplitude. Then by determining the highest loss amplitude, features can be obtained. A sample is finally described by 12 feature parameters.

8. Generation of audio model using Attention based convolutional capsule network (Attn_CCNet)

The difficult task of audio production is at the heart of several interesting topics, including voice conversion, music synthesis, and text-to-speech synthesis. The unique challenge with audio synthesis is that the effective semantic-level signal frequently has a substantially different dimensionality from the raw audio signal. Here Attn_CCNet is used for the generation of audio model. The architecture of Attn_CCNet is depicted in figure 5.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

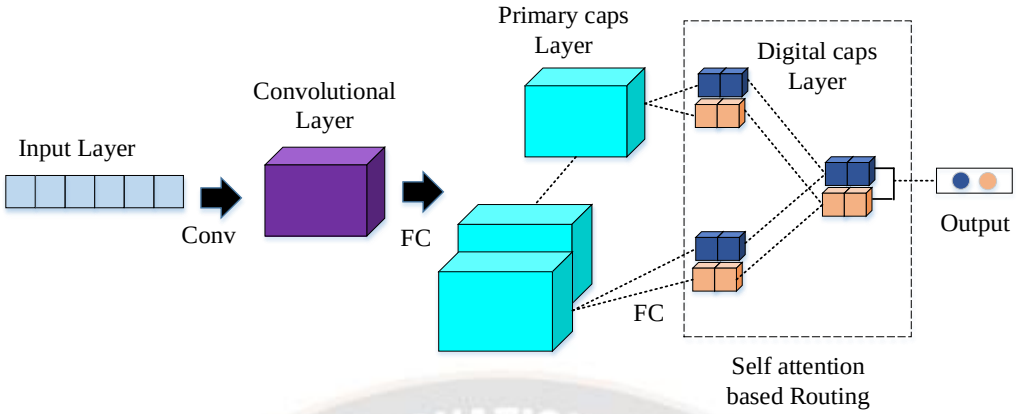


Figure 5: Architecture of Attn_CCNet

The feature extracted input is given to the convolutional layer. After the convolution operation, the feature mappings can be extracted. Each convolution kernel corresponds to a distinct piece of the feature. To extract the phrase's semantic features is the convolution layer's objective. Equation (15) performs the convolution on each sentence's matrix $A_{TT} = att_1, ...att_2, ...att_n$, it is the outcome of the word attention model.

$$S = F(W * A_{TT} + B) \quad (15)$$

Here B and W signifies the bias of network and weight matrix, correspondingly and nonlinear activation function is represented as $f = relu = Max(0, X)$. The convolutional operations extracted feature matrix is designated as S. The feature representation is aggregated and streamlined by the max pooling layer. To choose the top-K value of each filter to represent the semantic information, K-Max pooling is used. The dimensions of each convolution kernel's feature vector are noticeably smaller after the pooling procedure, and the most crucial semantic information is set aside.

Here, a capsule network layer was employed to determine the mapping of each input feature to the joint embedding space. A capsule is made up of an activation, which reflects the probability of an entity existing, and a multi-dimensional pose vector that represents the entity's attributes. The activations are formed by a linear layer followed by a sigmoid non-linearity, and a learnt linear layer derives a set of primary capsule postures from each input modality characteristic. These capsules are utilized by the self-attention-based routing system. The two components of Digit Caps entire connectivity layers under the management of Tanh and ReLU. By lowering the Euclidean distance between the output and training images, the sigmoid layer's output is produced. Digit Caps uses the appropriate label as a restoration target during training. Equation (16) displays the Caps Net S_j input.

$$S_j = \sum_i C_{ij} \hat{p}_{i/j} \quad (16)$$

The addition of entire prediction vectors from capsule of previous layer $\hat{p}_{i/j}$ makes up the overall value in a capsule S_j input. By dividing it by a weight matrix w_{ij}, the output of preceding layer $\hat{p}_i$ is determined as in equation (17).

$$\hat{p}_{i/j} = w_{ij} . \hat{p}_i \quad (17)$$

The output length of vector from the capsule layer signifies the likelihood that an entity signified through the capsule present in an input. Here, the squash function is scaled to 0.5 rather than 1. The squash function is utilized to normalize the output vector. The capsule layer's output vector indicates the probability of entity exists in input. A non-linear function is utilized to compress the short vector to a length nears to 0 and a lengthy vector nears to 1. Therefore, the squash function is scaled to 0.5 instead of 1. Consequently, the squashing equation is modified as in (18). It has been demonstrated that using this strategy, Caps Net's accuracy increased.

$$V_j = squash(S_j) = \frac{\| S_j \|^2}{1 + \| S_j \|^2} \Rightarrow \frac{\| S_j \|^2}{0.5 + \| S_j \|^2} \quad (18)$$

Despite the difficulties of solving by conventional CNNs using Caps Net, the thickness, shift, scale, etc. which define the object may be used to accurately recognise the item using capsules formed by the collection of neurons, location, angular value and change posture. Routing-by-agreement has been suggested as a method for learning the features.

Capsule networks transmit information from one layer to the next using routing. Similar mechanisms for determining agreement between high-dimensional vectors can be provided via self-attention. Effective syntactic or semantic features from the audio can be captured by Self-Attention. The output of convolutional layer at each time t is calculated using a single layer perceptual network in order to determine the audio's self-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

attention weight. A multi-head self-attention mechanism by connecting the primary capsules pose vectors to the pose feature space of the secondary capsules.

The features obtained by the generated models of text, image and audio are fused together by feature ensemble approach. From the fused feature vector, the optimal features are trained through IBRO algorithm.

F. Ensemble Feature Fusion Approach

For classification tasks, multimodal data is typically used to get excellent performance. Nevertheless, it might be difficult to combine different types of data from multiple modalities (multimodal data), particularly if one is interested in heterogeneous data. By deciding to display the finest of each modality, the fusion aims to correlate the components of each and improve the quality of what is exhibited. The features obtained by the generated models of text, image and audio are fused together by feature ensemble approach. Figure 6 depicts the ensemble feature fusion approach.

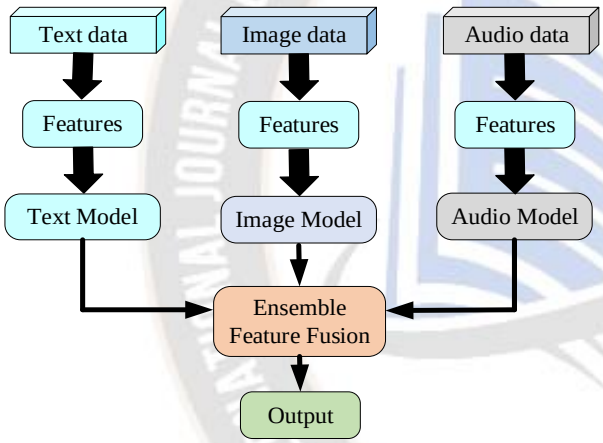


Figure 6: Ensemble feature fusion approach

The ensemble feature fusion is occurred in the stage of decision making. When the data sources significantly differ from one another in terms of sampling rate, data complexity, and unit of measurement, this methodology is much simpler than the early fusion method. This method frequently produces superior results since errors from various models are handled independently, resulting in uncorrelated errors. From the fused feature vector, the optimal features are trained through IBRO algorithm**.**

9. Optimal feature selection using Improved Battle royal optimization

Battle Royale games served as an inspiration for the BRO algorithm. In this type of fighting game, players engage in combat with one another in a harsh atmosphere, each trying to outlast the other players while also killing as many of them as they can. A player will re-spawned at a randomly selected area of the playing field if they sustain damage for a defined amount of time. Initial candidate solutions in the BRO method are randomly distributed around the problem space, just like in Battle Royale games.

In the IBRO, weight based distribution is used to select the features optimally. Afterward, each solution would be compared to its closest neighbour, and the one with the higher fitness value would be declared the victor and the inferior one the loser. Each candidate solution has a parameter that stores the damage (loss) level of each solution; this parameter is increased after each damage. A solution will be redistributed in accordance with equation (19) if it sustains damage repeatedly for a threshold period of time, which varies depending on the issue to be solved and ranges from (3) to (6). Equation (20) will handle the reallocation if its damage level falls below the threshold. In order to move closer to the best solution yet discovered, the candidate solution is reinitialized in each reallocation.

$$X^{i}_{d,dam} = w_t(Ub_d - Lb_d) + Lb_d \tag{19}$$

$$X^{i+1}_{d,dam} = X^{i}_{d,dam} + w_t(X_{best,d} - X^{i}_{d,dam}) \tag{20}$$

Here, the lower and upper boundaries of problem space's dimension d are denoted by the letters Lb_d and Ub_d, respectively. Thus the features are selected optimally by using this IBRO approach.

G. Classification using Convolutional duo Gated recurrent unit with auto encoder (C-Duo GRU_AE)

The model consists of dense and layers of output and input, embedding, and layers of convolutional layers, and layers that are merged with BiGRU and capsule networks. Figure 7 represents the structure of proposed HCov duo-caps framework.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

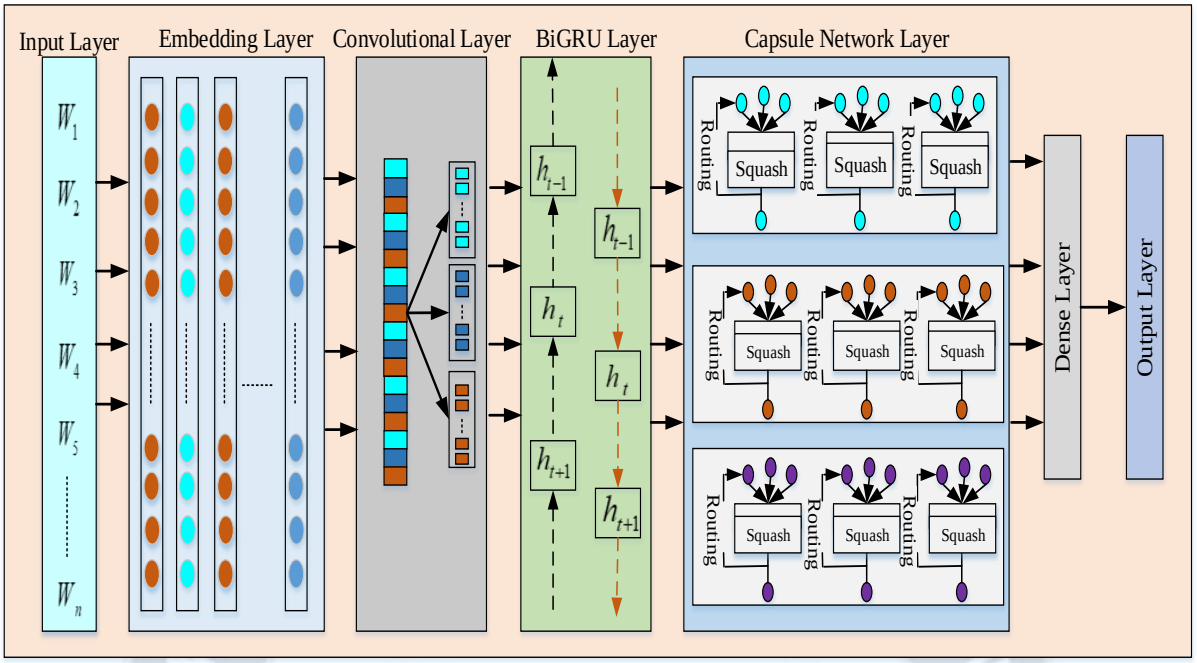


Figure 7: The HCov duo-Caps models structure

The network of input receives the pre-processed data from the HCov duo-Caps model and tokenizes it before translating every sign into a single integer utilizing the dictionary related index value. As a result, the layer transforms the provided text into a numerical value.

The embedding layer gains an ability to represent data into lows-dimensional dense vector through training on extensive real-world criteria. It recognises the connected data with comparable vector representations. Applications for the embedding layer range from recommender systems to machine translation and data classification.

The convolutional layer is applied by the HCov duo-caps over embedding vector to take the spatial properties. The convolutional layer employs 128 filters with 3 different sizes of filter to extricate temporal and spatial features associated with hate from the 128 sequences. The word window X_t of the nth feature sequence F_n is shown in equation (24), where the terms b, $F(.)$ and W_t denote the bias, filter weight, and ReLU, correspondingly. Equation (23) depicts the word window X_t of nth feature sequence F_n, where b, W_t, and $F(.)$ stand for the bias, filter weight, and ReLU correspondingly.

$$F_n = F(W_t.X_t + b) \qquad (21)$$

The 128 filters performs convolutional operation and extracts the sequence of features as $F_s = [F_1, F_2, ...F_{128}]$ from the input data. The max-pooling operation is used to acquire the underlying feature map. The output from each filter is concatenated by HCov duo-Caps to retrieve the final feature vector, which is used as an input to the subsequent layer.

To retrieve sequences from backward and forward directions in sequential modelling problems, a type of RNN known BiGRU is employed. BiGRU combines a backward and forward GRU to produce subsequent and previous feature sequences. In the presented HCov duo-Caps model, BiGRU layer is used to transform convolutional layer output into backward and forward sequences in which contextual information is added. The outputs from BiGRU are presented in forward and backward directions in Equations (24) and (25), correspondingly.

$$\vec{H}_f = \vec{G}_{RU}(L_{f_s}), \quad N \in [1,2...128] \qquad (22)$$

$$\overleftarrow{H}_b = \vec{G}_{RU}(L_{f_s}), \quad N \in [128...1,2] \qquad (23)$$

The output of BiGRU is a depiction of input text that includes hate by integrates the context of backward and forward directions. The BiGRU-based demonstration of the input text for a specific sequence of feature F_s is made up of combination of the backward $\overleftarrow{H}_b$ and forward $\vec{H}_f$ hidden states. In order to obtain hate, combining the sequence based contextual information, the two hidden states combine the data gathered surrounding L_{f_s}. Equation (26) finally depicts the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

concatenated sequence containing contextual information as a final state of hidden, H_t is sent to capsule network layer.

$$H_t = [\overleftarrow{H}_b, \overrightarrow{H}_f] \qquad (24)$$

Salient characteristics cannot be extracted using traditional CNN. Application of the pooling technique results in a loss of important information as well. By using Average pooling, Max, and Min, strategies, activation functions generate significant information that is lost when features are retrieved. A capsule can include other capsules in a network of capsules. Additionally, neurons are assembled into a capsule to extract semantic and syntactic data. When compared to conventional neural network models like CNN, the representation of capsule network stands further effective and comprehensive. As opposed to the CNN pooling layer, the capsule network produces vectors instead of scalar values. The capsule network layer receives the final hidden state H_t , which represents the BiGRU layer's output. Equations (27) and (29) are used to determine the resultant output of the capsule network. The non-linear activation in equation (27) is considered to transform final hidden state H_t of the BiGRU into a feature capsule U_i.

$$\hat{U}_{i/j} = w_{ij} U_i \qquad (25)$$

Here the input and output correlation are determined by U_i. The $\hat{U}_{i/j}$ denoted as the prediction vector and w_{ij} is the weight matrix.

The coupling coefficients C_{ij} are calculated using the dynamic routing method. The hate-related terms from the input data are disregarded by this method as being minor and irrelevant. Equation (29) is used to calculate a capsule's output S_j , which is the sum of all the prediction vectors. Here C_{ij} is the coupling coefficient.

$$S_j = \sum_{i=1}^{n} C_{ij} \hat{U}_{i/j} \qquad (26)$$

Equation (29) uses a squash function to normalise the final output vector V_j , it contains tokens from the input data in various orientations and local orders.

$$V_j = \frac{\| S_j \|^2}{1 + \| S_j \|^2} \frac{S_j}{\| S_j \|} \qquad (27)$$

The output vector V_j created by a capsule layer is delivered to fully connected layer. The events are subsequently categorised as protests, natural disasters, elections, sports, and festivals and public safety in the output layer. The binary cross-entropy loss function is engaged in suggested HCov duo-Caps model.

IV. RESULTS AND DISCUSSION

The proposed event detection in big data based on Optimization-based ensemble learning is implemented in python platform. Event dataset of Multi-modality (MMED) and Multi-domain are the source of the information used to estimate the efficiency of multimodal event categorization using the C-Duo GRU_AE classifier model. The efficiency of proposed scheme is assessed using the existing Decision tree (DT), LSTM, Naïve Bayes (NB), KNN and SVM in terms of Receiver Operating Characteristic Curve (ROC), F1 score, accuracy, precision, recall, and processing time. According to proposed structure, the input is divided into categories such as public safety, protest, natural disaster, election, sports, and festivals.

A. Dataset description

Multi-domain and Multi-modality (MMED) event Database [28] is comprised of textual news of 25,165 pieces were gathered from several online news media sources of data, such as the New York Times, NBC, Google, NBC, Fox, and Yahoo! to name a few. 4,473 Flickr members use social media to share a total of 76,516 Flickr image posts. All the data samples present in the dataset is related to the 412 real world events. To create the entire database as multi-modal data, the samples of audio are gathered from the platform of social media. 20% of the dataset is used for testing and 80% is used for training as a way to evaluate the results. The suggested classifier model uses the multimodal data as its input to classify various real-world events, such as protests, elections, natural disasters, sporting events, and festivals, effectively.

B. Performance metrics

In this section, effective performance measures like Recall, ROC, F1-score, Accuracy, Specificity, Processing Time, Sensitivity, and Precision are evaluated from the proposed technique.

1. Accuracy

Accuracy is determined by dividing the number of components $T_{positive}$ and $T_{Negative}$ by the sum of the components $T_{positive}$, $T_{Negative}$, $F_{positive}$, and $F_{Negative}$. The following equation can be used to determine accuracy.

$$\text{Accuracy} = \frac{T_{Positive} + T_{Negative}}{T_{Positive} + T_{Negative} + F_{Positive} + F_{Negative}} \qquad (28)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

$T_{Positive}$ Signifies the true positive, $T_{Negative}$ signifies the true negative, the $F_{Positive}$ denotes the false positive and $F_{Negative}$ denotes the false negative value.

2. Sensitivity

This is the overall amount of original positive data that is correctly classified.

$$\text{Sensitivity} = \frac{T_{positive}}{T_{Positive} + F_{Negative}} \quad (29)$$

3. Specificity

The ratio of $T_{Negative}$ to the total number of components that belong to the negative class (i.e., the total of $T_{Negative}$ and $F_{Positive}$) can be used to express it. Equation (33) displays the mathematical expression.

$$\text{Specificity} = \frac{T_{Negative}}{T_{Negative} + F_{Positive}} \quad (30)$$

4. Precision

Represented as a ratio of the total number of $T_{positive}$ to the total number of component tags, according to the positive class (i.e., the sum of $T_{positive}$ and $F_{positive}$).

The precision is indicated by the term Positive Predictive Value (PPV). The following are ways to gauge precision:

$$\text{Precision} = \frac{T_{Positive}}{T_{Positive} + F_{Positive}} \quad (31)$$

5. Recall

Represented as a ratio of the total number of $T_{positive}$ to the total number of component tags, according to the negative class (i.e., the sum of $T_{positive}$ and $F_{Negative}$).

$$\text{Recall} = \frac{T_{Positive}}{T_{Positive} + F_{Negative}} \quad (32)$$

6. F1- score

It is measured using the harmonic mean of recall and precision as shown in Equation (36).

$$\text{F1-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (33)$$

7. Processing time

It is the time taken to execute the proposed methodology. For the better performance the processing time should be minimum.

$$P_{time} = C_{time} - B_{time} \quad (34)$$

Here, P_{time} denotes the run time, C_{time} denotes the process completion time and B_{time} denotes the process beginning time.

8. ROC Curve

A graph displaying the effectiveness of a classification model at each classification threshold is known as a ROC curve. Two parameters are plotted on this curve:

The following definition of True Positive Rate (TPR), which is a synonym for recall:

$$T_{Positive\ Rate} = \frac{T_{Positive}}{T_{Positive} + F_{Negative}} \quad (35)$$

The definition of False Positive Rate (FPR) is as follows:

$$F_{Positive\ Rate} = \frac{F_{Positive}}{F_{Positive} + T_{Negative}} \quad (36)$$

C. Performance examination

In this section, performance of proposed scheme is validated with different current approaches. The whole evaluation outcome of suggested C-Duo GRU_AE -based event detection strategy is shown in table 2 below. The comparison validation of developed strategy in regards of accuracy, precision, Recall, F1 score and Training time is also stated.

TABLE 2: Comparison examination of proposed scheme

Techniques	Accuracy (%)	Recall (%)	F1-score (%)	Training time (s)	Precision (%)
DT	99.70	99.71	99.72	36	99.71
NB	53.86	52.44	68.10	40	52.24
LSTM	99.72	99.69	99.72	35	99.75
KNN	85.22	85.72	85.38	24	85.04
SVM	60.67	60.12	71.91	18	56.15
EEDEM	99.87	99.87	99.87	60	99.87
Proposed	**99.93**	**99.93**	**99.91**	**17**	**99.92**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

As shown in table 2, the suggested method is assessed utilizing four statistical metrics: precision, accuracy, recall, F1-score, and training duration. In general, accuracy value describes how closely the derived value matches the true value. The achieved classification accuracy is 99.93%, showing that the suggested strategy to event detection is quite effective. Additionally, the precision of this method for event detection is roughly 99.92%. Similarly, the recall, f1-score values, and training time obtained are 99.93%, 99.91%, and 17 correspondingly. The validation of output in regards of accuracy is indicated in figure 8.

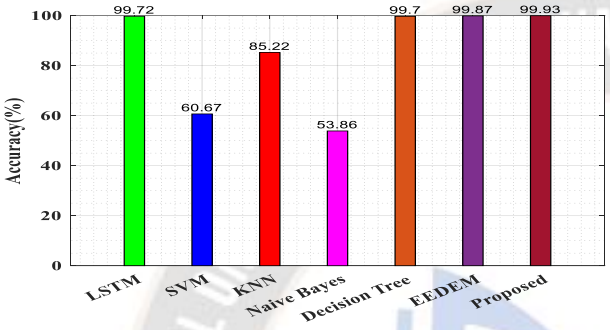
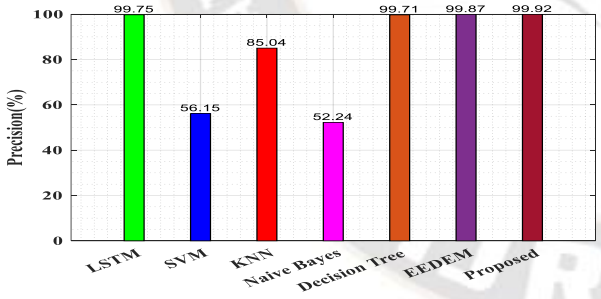


Figure 8: Performance analysis of accuracy

In figure 8, presented scheme achieves improved accuracy 99.93% value than other current Decision Tree (DT), NB (Naive Bayes), LSTM (Long short term memory), KNN (K-nearest neighbour's algorithm), EEDEM (Emergency event detection ensemble model) and SVM techniques. The precision validation of existing and proposed method is illustrated in figure 9.



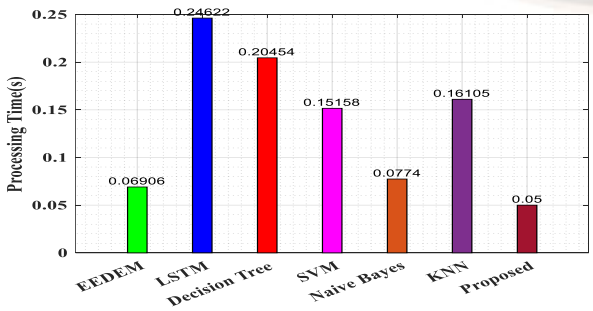


Figure 12: Performance analysis of processing time

Figure 9: Comparison examination of precision

The figure 9 depicts that the precision outcome of developed scheme is 99.92%. Reaches substantial performance than other existing DT, NB, LSTM, KNN, EEDEM and SVM techniques. Furthermore, the recall performance is portrayed in figure 10.

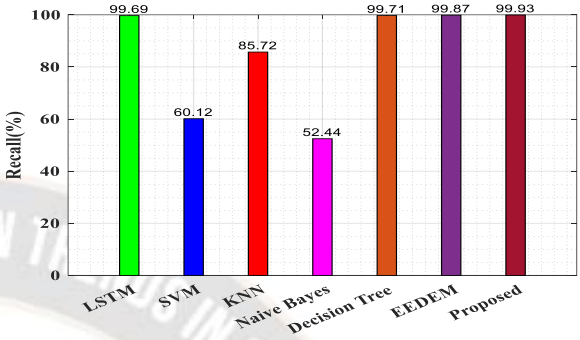


Figure 10: Comparison examination of recall

The figure 10 proves that the recall outcome of presented approach is 99.93%, achieves substantial outcome than current DT, NB, LSTM, KNN, EEDEM and SVM techniques. Performance computation of F1-score in introduced approach compared to previous work is illustrated in figure 11.

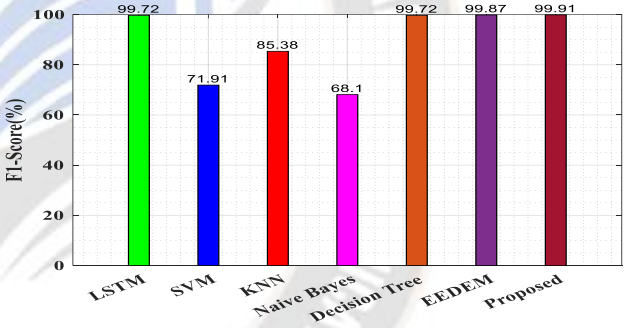


Figure 11: Comparison examination of F1-score

The figure 11 depicts that F1 score outcome of developed approach is 99.91%, which obtains good improvement than other current approaches. Performance validation in regards of processing time of presented scheme comparing with other work is shown in figure 12.

Figure 12 demonstrates that processing time comparison of presented approach to current DT, NB, LSTM, KNN, and EEDEM SVM methodology. The processing time of introduced scheme takes 0.05 second which is more highly consistent. Figure 13 depicts the training time of proposed method.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

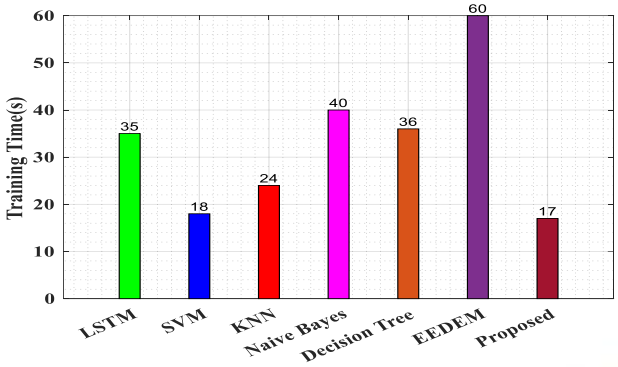


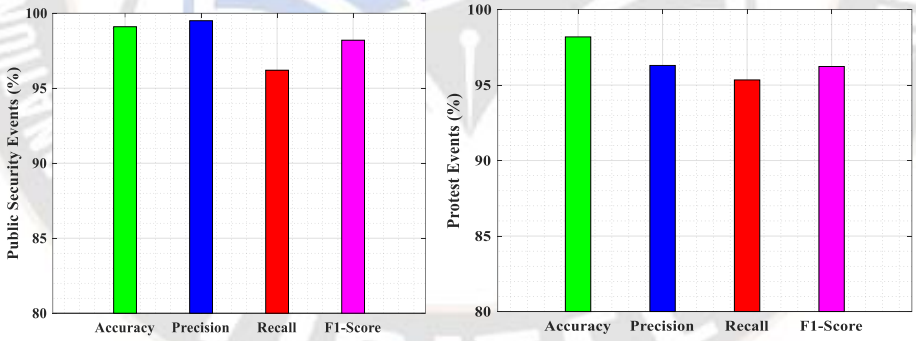
Figure 13: Performance analysis of training time

Figure 13 demonstrates that proposed training time comparison to other existing DT, NB, LSTM, KNN, EEDEM and SVM schemes. The training time of presented approach takes 17 second, which is much lesser than other compared techniques. Table 3 depicts the performance of different events.

TABLE 3: Performance of various events

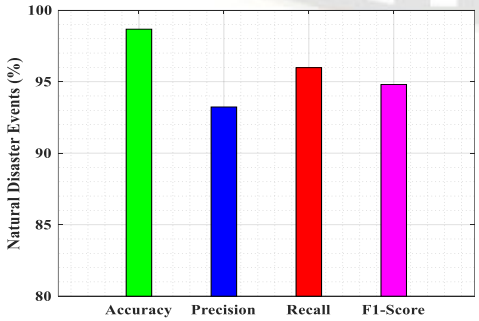
Events	Accuracy	Precision	Recall	F1-score
Public security	99.1	99.5	96.2	98.2
Protest	98.19	96.3	95.34	96.23
Natural disaster	98.67	93.23	95.98	94.8
Election	98.6	96.45	98.34	97.23
Sports	99.2	93.45	94.45	95.23
Festival	98.2	89.3	89.34	90.34

The class-by-class comparison of the suggested strategy is shown in table 3. The proposed strategy is evaluated in Table 3 based on five distinct types of events, including public safety, protests, natural disasters, elections, sports, and festivals. It is clearly depicted that sports event has higher accuracy of 99.2%. The performance in terms of accuracy of public security, natural disaster, protest, election and festival is 98.2%, 96.23%, 94.8%, 97.23% and 90.34% respectively. The performance in terms of precision the public security has the highest value of precision of 99.5. The Protest, Sports, Natural disaster, Election, and Festival has the precision value of 96.3%, 93.23%, 96.45%, 93.45%, and 89.3% respectively. The higher recall value obtained by the election event. The Protest, Natural disaster, Public security, Sports, and Festival has the recall value of 96.2%, 95.98%, 95.34%, 94.45%, and 89.34 respectively. The performance in terms of F1-score public security acquired the highest value. The Protest, Natural disaster, Election, Sports, and Festival has the F1-score of 96.23%, 94.8%, 97.23%, 95.23%, and 90.34% respectively. The comparison analysis of various events are depicted in figure 14.

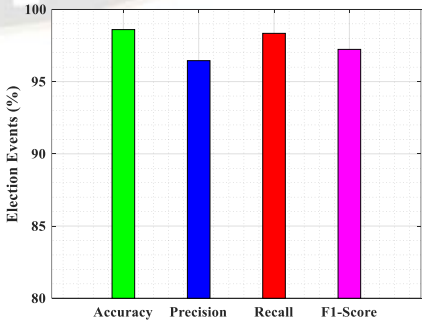


(a)

(b)

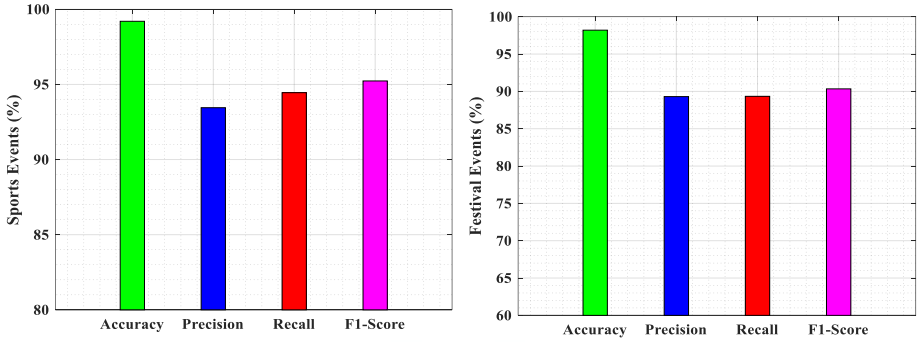


(c)



(d)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*



(e) (f)

Figure 14: Performance comparison of various events (a) Public security, (b) protest, (c) natural disaster, (d) election, (e) sports and (g) festival

Figure 14 denotes the comparison analysis of different events such as Public security, natural disaster, election, sports, protest, and festival. Here, the performance of proposed method on each event is provided. In an analysis of each individual event, these performance results are better. The attained accuracy of Public security, natural disaster, protest, election, sports and festival are 99.1%, 98.19%, 98.67%, 98.6%, 99.2%, and 98.2% correspondingly. The precision outcome of consistent events are 99.5%, 93.23%, 96.3%, 96.45%, and 93.45% correspondingly. Similarly the recall values are 96.2%, 95.34%, 95.98%, 98.34%, 94.45%, and 89.34% correspondingly. The f1-score outcome is much improved and the outcomes are 98.2%, 96.23%, 94.8%, 97.23%, 95.23%, 90.34% correspondingly. Moreover, the performance of presented methodology compared with various current works in terms of precision, accuracy, recall and F-score. The performance analysis of accuracy is shown in figure 15.

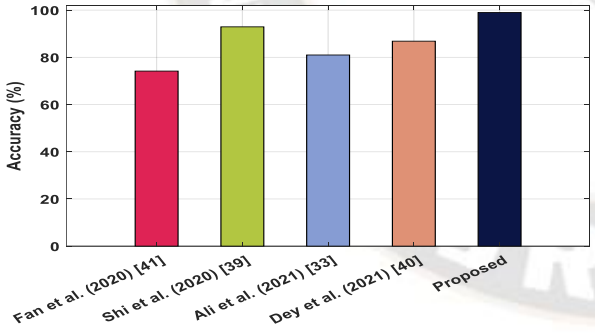


Figure 15: Performance analysis of accuracy

In figure 15, accuracy (99.93%) comparison with different existing approaches [29] are given. Here, the accuracy examination is done with different existing event detection works such as Fan et al. (2020) [40], Shi et al. (2020) [38], Ali et al. (2021) [32], and Dey et al. (2021) [39]. This proved that the proposed methodology attains good accuracy performance in multimodal event detection than other existing works. Subsequently, the comparison validation of recall performance is shown in figure 16.

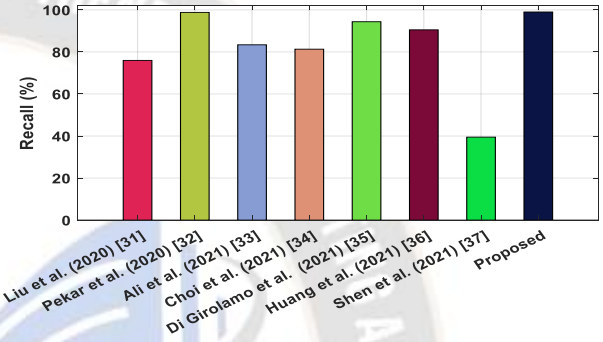


Figure 16: Performance comparison on recall

In figure 16, the recall performance is analysed with different existing schemes [29]. The proposed scheme attains higher recall (99.93%) outcome than compared works. This clearly illustrates that presented work provides improvements than existing schemes. Furthermore, the comparison on F-score is depicted in figure 17.

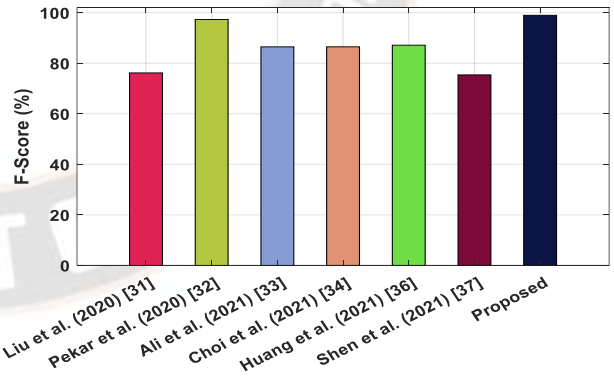


Figure 17: Performance comparison on F-score

The F-score comparison given in figure 17 depicts that the performance enhancement than the compared works. The proposed work attains high F-score (99.91%) performance than other exiting works. This showed the efficiency of presented scheme in event detection. Furthermore, the precision outcome is given in figure 18.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

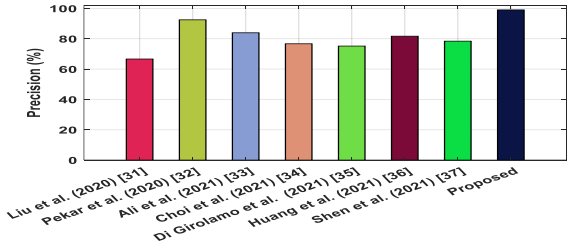


Figure 18: Performance validation of Precision

In figure 18, proposed scheme is compared with various current methods in regards of precision. This showed that introduced scheme achieves high precision performance (99.93%) than compared existing schemes in Liu et al. (2020) [30], Pekar et al. (2020) [31], Ali et al. (2021) [32], choi et al. (2021) [33], Di Girolamo et al. (2021) [34], huang et al. (2021) [35], Shen et al. (2021) [36]. The comparison analysis on accuracy achieves higher outcome than compared existing schemes. Furthermore, the comparison on AUC is given in figure 19.

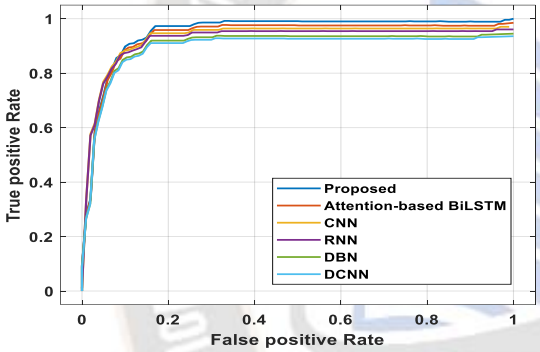


Figure 19: Performance comparison on AUC

In figure 19, the proposed methodology is compared with different current strategies in regards of AUC. This showed that the presented scheme obtains good performance than compared existing schemes like deep convolutional neural network (DCNN), DBN, Recurrent neural network (RNN), CNN, and Attention based BiLSTM [37]. Thus, the proposed scheme clearly demonstrates that presented scheme obtains good performance than the compared schemes.

V. CONCLUSION

Event detection is regarded as a crucial process in a wide range of industries. There are several research studies in this area because of its wide range of applications. Maximum event detection methods uses one specific category of information (i.e. text or image or audio). As a result, it is challenging to evaluate or demonstrate the effectiveness of the detection procedure. Therefore, this work proposed a novel text, image, and audio-based event detection method. The proposed work includes pre-processing for text, images, and audio. Furthermore, different feature extraction approaches are considered to extract the features. Afterwards, model generation is performed for image, text and audio data individually. Finally, event detection is performed for the classification of various events separately. Here, the C-Duo GRU AE neural network structure is used in the classification stage to produce better results. The proposed methodology detects the multi-modal events accurately. Furthermore, the effectiveness of the presented approach is estimated using a number of existing approaches. The proposed technology achieves improved performances in terms of Accuracy (99.93%), Precision (99.93%), Recall (99.93%), F1-score (99.91%) processing time (17sec) and training time (0.05sec). This indicates that the suggested methodology achieves superior results to the compared approaches. In future, the proposed framework will be extended with improved approaches. Moreover, more recent datasets can be incorporated to validate the event detection performance. Also, more performance metrics can be considered to validate the performance of the work.

ACKNOWLEDGMENT

None

REFERENCES

[1] S. Wang, G. Yu, Z. Cai, X. Liu, E. Zhu, J. Yin et alVideo Abnormal Event Detection by Learning to Complete Visual Cloze Tests", 2021, arXiv preprint arXiv:2108.02356.

[2] A. Hodorog, I. Petri, Y. Rezgui, "Machine learning and Natural Language Processing of social media data for event detection in smart cities". Sustainable Cities and Society, pp. 104026, vol. 85, 2022.

[3] Y. Zhang, C. Ridings, A. Semenov, "What to post? Understanding engagement cultivation in microblogging with big data-driven theory building". International Journal of Information Management, vol. 102509, 2022.

[4] M. Mężyk, M. Chamarczuk, M. Malinowski, "Automatic image-based event detection for large-N seismic arrays using a convolutional neural network". Remote Sensing, pp. 389, vol. 13, no. 3, 2021.

[5] E.A. Hinojosa-Palafox, OM Rodríguez-Elías, JA,Hoyo-Montaño JH Pacheco-Ramírez, JM Nieto-Jalil. "An analytics environment architecture for industrial cyber-physical systems big data solutions". Sensors, pp. 4282, vol. 21, no. 13, 2021.

[6] Y. Cao, H. Peng, J. Wu, Y. Dou, J. Li, P.S. Yu, "Knowledge-preserving incremental social event detection via heterogeneous gnns". In Proceedings of the Web Conference pp. 3383-3395, vol. 2021, 2021, April.

[7] A. Hodorog, I. Petri, Y. Rezgui, "Machine learning and Natural Language Processing of social media data for event detection in smart cities". Sustainable Cities and Society, pp. 104026, vol. 85, 2022.

[8] Y. Chen, Y. Li, Z. Wang, A.J. Quintero, C. Yang, W. Ji, "Rapid perception of public opinion in emergency events through social media". Nat. Hazards Rev, pp. 04021066, vol. 23, no. 2, 2022.

[9] F. Agneessens, G.J. Labianca, "Collecting survey-based social network information in work organizations". Social Networks, pp. 31-47, vol. 68, 2022.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

[10] Y. Wu, H. Huang, N. Wu, Y. Wang, M.Z. Bhuiyan, T. Wang, "An incentive-based protection and recovery strategy for secure big data in social networks". Information Sciences, pp. 79-91, vol. 508, 2020.

[11] Morzelona, R. (2021). Human Visual System Quality Assessment in The Images Using the IQA Model Integrated with Automated Machine Learning Model . Machine Learning Applications in Engineering Education and Management, 1(1), 13–18. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/5

[12] H.U. Khan, S. Nasir, K. Nasim, D. Shabbir, A. Mahmood, "Twitter trends: A ranking algorithm analysis on real time data". Expert Systems with Applications, pp. 113990, vol. 164, 2021.

[13] M. Paul, M. Danelljan, L. Van Gool, R. Timofte, "Local memory attention for fast video semantic segmentation". In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) pp. 1102-1109, 2021, January.

[14] Z Wu, W Wang, Y Peng "Deep learning-based UAV detection in the low altitude clutter background", 2022. arXiv preprint arXiv:2202.12053.

[15] E. Curry, D. Salwala, P. Dhingra, F.A. Pontes, P. Yadav, "Multimodal event processing: A neural-symbolic paradigm for the internet of multimedia things". IEEE Internet of Things Journal 2022.

[16] S. Hussain, M. Mubeen, A. Ahmad, S. Fahad, W. Nasim, H.M. Hammad, G.M. Shah, B. Murtaza, M. Tahir, S. Parveen, "Using space–time scan statistic for studying the effects of COVID-19 in Punjab, Pakistan: A guideline for policy measures in regional agriculture". Environmental Science and Pollution Research, pp. 1-14, 2021.

[17] C. Choi, S.Y. Hong, "MDST-DBSCAN: A Density-Based Clustering Method for Multidimensional Spatiotemporal Data". ISPRS International Journal of Geo-Information, pp. 391, vol. 10, no. 6, 2021.

[18] M. Elbadawi, S. Gaisford, A.W. Basit, "Advanced machine-learning techniques in drug discovery". Drug Discovery Today, vol. 26, no. 3, 2021, pp. 769-777.

[19] Vyas, A. ., & Sharma, D. A. . (2020). Deep Learning-Based Mango Leaf Detection by Pre-Processing and Segmentation Techniques. Research Journal of Computer Systems and Engineering, 1(1), 11–16. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/18

[20] N.S. Nafis, S. Awang, "An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification". IEEE Access, pp. 52177-52192, vol. 9, 2021.

[21] Z. Huo, W. Zhu, P. Pei, "Network Traffic Statistics Method for Resource-Constrained Industrial Project Group Scheduling under Big Data". Wireless Communications and Mobile Computing, vol. 2021, 2021.

[22] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches". Transactions on Emerging Telecommunications Technologies, pp. e4150, vol. 32, no. 1, 2021. 21. E. Alomari, I. Katib, Aiiad Albeshri, Tan Yigitcanlar, and Rashid Mehmood, "Iktishaf+: a big data tool with automatic labeling for road traffic social sensing and event detection using distributed machine learning". Sensors, pp. 2993, vol. 21, no. 9, 2021.

[23] K. Alfalqi, M. Bellaiche, "An Emergency Event Detection Ensemble Model Based on Big Data". Big Data and Cognitive Computing, pp. 42, vol. 6, no. 2, 2022.

[24] E. Alomari, I. Katib, R. Mehmood, "Iktishaf: A big data road-traffic event detection tool using Twitter and spark machin learning". Mobile Networks and Applications, pp. 1-16, 2020.

[25] Y. George, S. Karunasekera, A. Harwood, K. H. Lim, "Real-time spatio-temporal event detection on geotagged social media". Journal of Big Data, pp. 1-28, vol. 8, no. 1, 2021.

[26] H. M. Lee, S.J. Lee, "A Study on Security Event Detection in ESM Using Big Data and Deep Learning". International Journal of Internet, Broadcasting and Communication, pp. 42-49, vol. 13, no. 3, 2021.

[27] K. Alfalqi, M. Bellaiche, "An Emergency Event Detection Ensemble Model Based on Big Data." Big Data and Cognitive Computing pp. 42, vol. 6, no. 2, 2022.

[28] A. D. Smith, "Event detection in educational records: an application of big data approaches." International Journal of Business and Systems Research pp. 271-291, vol. 15, no. 3, 2021.

[29] Mr. Ather Parvez Abdul Khalil. (2012). Healthcare System through Wireless Body Area Networks (WBAN) using Telosb Motes. International Journal of New Practices in Management and Engineering, 1(02), 01 - 07. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/4

[30] Z. Yang, Z. Lin, L. Guo, Q. Li, W. Liu, "MMED: a multi-domain and multi-modality event dataset." Information Processing & Management pp. 102315, vol. 57, no. 6, 2020.

[31] Brown, R., Brown, J., Rodriguez, C., Garcia, J., & Herrera, J. Predictive Analytics for Effective Resource Allocation in Engineering Education. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/91

[32] M. S. Mredula, N. Dey, M. S. Rahman, I. Mahmud, Y.Z. Cho, "A Review on the Trends in Event Detection by Analyzing Social Media Platforms' Data." Sensors pp. 4531, vol. 22, no. 12, 2022.

[33] Y. Liu, H. Peng, J. Li, Y. Song, X. Li, "Event detection and evolution in multi-lingual social streams." Frontiers of Computer Science pp. 1-15, vol. 14, 2020.

[34] V. Pekar, J. Binner, H. Najafi, C. Hale, V. Schmidt, "Early detection of heterogeneous disaster events using social media." Journal of the Association for Information Science and Technology pp. 43-54, vol. 71, no. 1, 2020.

[35] Ali, Daler, Malik Muhammad Saad Missen, and Mujtaba Husnain. "Multiclass event classification from text." Scientific Programming pp. 1-15, vol. 2021, 2021.

[36] D. Choi, S. Park, D. Ham, H. Lim, K. Bok, J. Yoo, "Local event detection scheme by analyzing relevant documents in social networks." Applied Sciences pp. 577, vol. 11, no. 2, 2021.

[37] R. Di Girolamo, C. Esposito, V. Moscato, G. Sperlí, "Evolutionary game theoretical on-line event detection over

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6966
***Article Received**: 24 March 2023 **Revised**: 20 May 2023 **Accepted**: 02 June 2023*

tweet streams." Knowledge-Based Systems pp. 106563, vol. 211, 2021.

[38] L. Huang, G. Liu, T. Chen, H. Yuan, P. Shi, Y. Miao, "Similarity-based emergency event detection in social media." Journal of Safety Science and Resilience pp. 11-19, vol. 2, no. 1, 2021.

[39] C. Shen, Z. Li, Y. Chu, Z. Zhao, "GAR: Graph adversarial representation for adverse drug event detection on Twitter." Applied Soft Computing pp. 107324, vol. 106, 2021.

[40] K. Swapnika, D. Vasumathi, "Multimodal event detection in big data using multi-level fusion classifier". Indian Journal of Computer Science and Engineering (IJCSE), pp. 796-811, vol. 13, no. 3, 2022.

[41] L. l. Shi, L. Liu, Y. Wu, L. Jiang, A. Ayorinde, "Event detection and multi-source propagation for online social network management." Journal of Network and Systems Management pp. 1-20, vol. 28, 2020.

[42] N. Dey, M. S. Rahman, M. S. Mredula, A.S.M. S. Hosen, I.-Ho Ra, "Using machine learning to detect events on the basis of Bengali and banglish Facebook posts." Electronics pp. 2367, vol. 10, no. 19, 2021.

[43] C. Fan, F. Wu, A. Mostafavi, "A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters." IEEE Access pp. 10478-10490, vol. 8, 2020.

[44] Solanki, S. ., Singh, U. P. ., Chouhan, S. S. ., & Jain, S. . (2023). Brain Tumour Detection and Classification by using Deep Learning Classifier. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 279 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2624

[45] Wang Wei, Natural Language Processing Techniques for Sentiment Analysis in Social Media , Machine Learning Applications Conference Proceedings, Vol 1 2021.