RESEARCH ARTICLE

GenVis: Visualizing Genre Detection in Movie Trailers for Enhanced Understanding

FAHEEM SHAUKAT[1], **NAVEED EJAZ**[2], **ZEESHAN ASHRAF**[3], **(Senior Member, IEEE)**,
MRIM M. ALNFIAI[4], **MONA MOHAMMED ALNAHARI**[4], **AND REEMIAH MUNEER ALOTAIBI**[5]

[1]Department of Computing and Technology, IQRA University, Islamabad Campus, Islamabad 44000, Pakistan
[2]School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada
[3]Department of Computer Science, The University of Chenab, Gujrat 50700, Pakistan
[4]Department of Information Technology, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[5]Information Technology Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, Saudi Arabia

Corresponding author: Zeeshan Ashraf (zeeshan.ashraf@ieee.org)

ABSTRACT Automatic movie genre detection is vital for improving content recommendations, user experiences, and organization. Multi-label generation detection assigns multiple labels to a movie and recognizes a movie's diverse themes. Although there are many existing methods for generating multiple genre labels from movies but do not provide comprehensive analysis and visual depiction. This work introduces GenVis, a visualization system that provides a better understanding of multi-label genres extracted from movie trailers. The system initially uses text and visual features to classify trailers and assign multiple genre labels and probabilities. Next, GenVis provides four visualization views: a video view for trailer observation, an overall genre view for getting insights into genre distribution, a genre timeline view for temporal genre evolution, and finally, a genre flow summary for more focused genre analysis. The system allows users to pause the frames, sort the results, and process multiple videos. The multi-label classification is rigorously evaluated using MSE, cross-entropy loss, precision, recall, F1-score metrics, achieving high accuracy, and demonstrating strong genre correlations with notable precision in effectively classifying and distinguishing movie genres. Additionally, a user evaluation for visualization evaluation demonstrated the effectiveness and intuitive usability of GenVis with a high overall rating of 4.25 out of 5.0.

INDEX TERMS Genre detection, movie trailers, video visualization, data visualization, video analysis.

I. INTRODUCTION

In a movie, genres are distinct categories that characterize the thematic content of a movie. Automatic genre detection of movies has gained significant popularity in recent years [1]. Automatically identifying movie genre(s) is vital for effective content recommendation, personalized user experiences, and content organization [2].

Multi-label genre detection in movies refers to assigning multiple genre labels to a single movie based on its content and characteristics. Unlike single-label classification, where a movie is assigned only one genre label, multi-label genre

The associate editor coordinating the review of this manuscript and approving it for publication was Qiang Yang.

detection considers that movies can often simultaneously belong to more than one genre [3]. This approach recognizes the complexity and diversity of movies that often blend various themes and elements.

The existing single and multiple-label genre detection datasets and methods only provide the final answer(s) in the form of the genre(s). The widely used datasets, including the IMDB dataset [4], [5], [6], [7], only provide labels of the genres. The methods do not provide intricate details of analysis that have been used to reach the answer in the form of genres. Moreover, the users cannot see the proportion of a particular genre in a movie [8]. The users may still struggle to make informed decisions about watching a movie when presented with multiple genre labels. Without visual

representation, the users might not have a clear sense of the prominence of each genre in the movie, making it harder for them to gauge whether a movie aligns with their preferences. Movies with a high number of genres might have their content misrepresented. A multi-label classification without visualizations could label a movie with many genres, even if some are only briefly present.

In this context, this work proposes a visualization system called "GenVis" to analyze the multi-label genres of movies extracted from the trailers. The results of multi-label genres can be visualized using four views: the video view, which provides users with the facility to watch movie trailers during analysis; the overall genre view, which shows the overall distribution of genre percentages in the trailer; the genre timeline view, which provides the users with a summary of genre evolution over time, and the genre flow summary, which allows to focus on a particular genre in more details related to time progression of genres. The system also enables users to pause and analyze a particular frame. Moreover, the results can also be sorted based on a particular genre. Lastly, the users can simultaneously process and visualize the overall genre distribution of up to 5 videos.

A dataset of 25 movie trailers is developed to experiment with the proposed method. Three annotators assign the probability of each genre after watching the trailer to generate the ground truth. The system is restricted to six genres: Adventure, Action, Romance, Family, Fantasy, and Horror. A multi-modal deep learning-based method using text and visual features classifies a movie trailer based on text and visual clues by assigning multiple genre labels and respective probabilities.

The accuracy of our multi-label genre classification is validated against user annotations through mean square error (MSE), cross-entropy loss, precision, recall, as well as F1-score metrics. It achieves a notably low MSE of 0.019, demonstrating high alignment with actual genre probabilities alongside a cross-entropy loss of 1.32, underscoring its precision in reflecting human genre perceptions. Precision, recall, along with F1 scores are calculated for each genre, with results detailed from the top 1 (when one genre is selected) to the top 5 (when the top five genres are selected) genre selections. Additionally, a co-occurrence matrix for the top three genres highlights strong correlations, especially between Action, Adventure, and Fantasy, identifying common themes while illustrating the complexity of distinguishing closely related genres such as Romance and Family. Moreover, a user study was conducted to evaluate the visualizations presented in GenVis. Involving 43 undergraduate computer science students, our user study assessed the effectiveness of the visualization system, revealing its comprehensibility and intuitive navigation with an overall positive rating of 4.25, highlighting its potential as a valuable tool for enhancing genre understanding. The outcomes of these evaluations demonstrate encouraging achievements in multi-label genre detection, highlighting the system's proficiency in accurately categorizing genres within movie trailers. Additionally, the

feedback from users regarding the visualization experience was predominantly positive, which reflects the system's ability to present intricate genre distribution patterns and temporal evolutions effectively.

The following are the major contributions of the paper:

1) The paper introduces an innovative visualization system called GenVis designed to analyze multi-label movie genres extracted from trailers. GenVis offers multiple views, including the video view, overall genre view, genre timeline view, and genre flow summary, providing users with comprehensive tools for exploring and understanding movie genre dynamics.

2) GenVis allows users to delve directly into movie trailers through the video view, offering an effortless experience for understanding genre representation. Users can watch trailers during analysis, gaining a holistic interpretation of genre dynamics present within each trailer.

3) GenVis provides a valuable temporal perspective through the genre timeline view. This view summarizes how genres develop and shift within the duration of a trailer. It aids in identifying patterns and trends, contributing to a richer contextual understanding of genre representation.

4) GenVis allows for concurrent processing and visualization of genre distributions for multiple videos, expediting comparative analysis. Users can make meaningful cross-trailer assessments and draw insightful conclusions regarding genre dynamics.

The rest of this paper is arranged as follows. Section II presents related work on movie genre detection and video visualization. Section III describes the dataset generation and the methodology used for multi-label genre classification. Section IV discusses the visualization design, while Section V discusses experiments and results. Finally, Section VI concludes the paper.

II. RELATED WORK

The literature has explored diverse video modalities for genre classification, encompassing movie posters, visual features from frames or shots, audio, textual content, and multi-modal features. This section summarizes recent research on multi-label genre classification using movie trailers, posters, textual attributes, and multi-modal features. A brief review of existing video visualization schemes has also been provided.

A. TRAILER BASED SCHEMES

The authors in [9] proposed a deep learning architecture in which ImageNet and Kinetics learn the representations. The transferability was improved by performing an adaptation stage that segments the input trailer based on shot transitions. Moreover, a dataset of 12,000 trailers was also proposed. Reference [10] proposed an attention-based spatiotemporal sequential framework that captured the local and global high-level sequential semantic features within the movie storyline. A sequential model with attention mechanisms

was developed to capture overarching sequential semantic attributes and emphasize specific elements within the movie's plot. Reference [11] proposed a deep affect-based framework for multi-label classification of genres in movies. The research introduced an approach for classifying movie genres based on emotions by using an ensemble of Inception V4, Bi-LSTM, and LSTM layers to generate distinct and comprehensive high-level attributes. In the work by Simoes et al. [12], a VGG-like 2D ConvNet structure was employed, starting with individual frame classification and subsequently employing diverse aggregation methods to derive an overall prediction for the trailer. Reference [13] utilized a ResNet architecture pre-trained on ImageNet and Places-360. This method harnessed frame-level representations with the CTT module for trailer classification. Some older methods employed hand-crafted features for multi-label genre classification from trailers [14], [15].

B. POSTER-BASED SCHEMES
Unal et al. [8] explored transfer learning for multi-label genre classification using six pre-trained movie poster models. Using an iterative stratification technique, they divided the dataset and fine-tuned models with added dense layers. Reference [16] conducted multi-label genre classification on posters using Gram layer in CNNs. ResNet was used as the base architecture, examining ResNet 18, 34, 50, and 152 on IMDB movie posters from 1913 to 2019 across 12 genres. Kundalia et al. [3] introduced a deep model employing transfer learning to forecast movie genres based on poster images. They trained a pre-trained Inception-V3 model for transfer learning to categorize 12 genres using IMDB dataset images. Barney and Kaya [17] introduced a movie genre detection study based on K-nearest neighbours, their custom deep architecture, and ResNet-34 for classification across five genres.

C. TEXT-BASED SCHEMES
Some schemes used scripts, subtitles, or movie plots for the task. The study [18] focused on predicting movie genres utilizing untapped subtitles by identifying high-frequency genre-specific words as features for machine learning models. Experiments on 964 English movies across six genres revealed optimal results with K-Nearest Neighbour (kNN) using 200 features, achieving 77.7% average precision. Reference [19] explored multi-label film genre classification using synopses, particularly for Portuguese films. They introduced the P-TMDb dataset, containing 13,394 Portuguese film synopses, and assessed genre classification with nine textual features and four multi-label algorithms. The study [20] used movie plot summaries. This study employed bidirectional LSTM (Bi-LSTM) for movie genre classification based on plot summaries. The approach involved segmenting each summary into sentences, associating genres with sentences, and training Bi-LSTM networks on sentence word representations.

D. MULTI-MODALITIES-BASED SCHEMES
A lot of recent methods have focused on using multi-modalities for genre classification. Reference [21] presented a multimodal movie genre classification approach using deep learning techniques, integrating diverse data sources, including audio, posters, plot summaries, and video frames. Through strategic decision-level and intermediate fusion, combining methods like concatenation and element-wise sum, they capitalize on the synergies between modalities to enhance classification accuracy. Reference [22] used various features, including Mel Frequency Cepstral Coefficients, Statistical Spectrum Descriptor, Local Binary Pattern with spectrograms, Long-Short Term Memory, and Convolutional Neural Networks. Bi et al. [23] introduced the Shot-based Hybrid Fusion Network, which combined single-modal feature fusion networks for both video and audio alongside a shot-based multi-modal feature fusion network. A late fusion element was integrated to make decisions at the video level. Braz et al. [24] worked on movie genre classification using movie posters and plots. A deep convolutional neural network with an embedded attention mechanism was used to analyze posters. The BERT model [25] was refined to analyze the textual context. Arevalo et al. [26] introduced a multimodal learning approach utilizing gated neural networks. Their model introduced a gated multimodal unit (GMC) that combines the movie poster and plot to create an intermediate representation through multimodal fusion techniques.

E. VIDEO VISUALIZATION
Data Visualization has been used extensively in various fields [27]. Video visualization uses visual methods to provide users with a quick overview of video content. Important information, like key features and important events, is extracted from the original video and presented using visual techniques. This makes it easier to recognize specific patterns in the video. Daniel and Chen first introduced video visualization [28]. Video visualization has been successfully applied in many applications such as sports video analysis [29], [30], [31], emotion analysis [32], surveillance [33], and biology applications [34]. However, no published work on visualizing movie genres exists to our knowledge.

III. MULTIMODAL GENRE CLASSIFICATION
This section describes the data preparation and the methodology used for the multi-label classification of trailers into genres. Let X be the set of movie trailers, where each trailer $x \in X$ is represented by a feature vector of dimension d. These features can include visual, audio, and textual embeddings extracted from the trailers. Let $G = \{g_1, g_2, \ldots, g_N\}$ be the set of possible genres, where N is the total number of genres. We define $Y = \{y_1, y_2, \ldots, y_N\}$ as the binary label vector representing the presence or absence of each genre in a trailer. If a trailer belongs to genre g_i, then $y_i = 1$; otherwise, $y_i = 0$. The goal is to learn

TABLE 1. List of movies in the dataset and their release years.

Movie	Release Year
Harry Potter and the Deathly Hallows: Part 1	2010
Clash of the Titans	2010
The Sorcerer's Apprentice	2010
The Twilight Saga: Eclipse	2010
Red Riding Hood	2011
I Am Number Four	2011
Snow White and the Huntsman	2012
Dark Shadows	2012
Warm Bodies	2013
Hansel & Gretel: Witch Hunters	2013
Dracula Untold	2014
Maleficent	2014
Cinderella	2015
Goosebumps	2015
The Jungle Book	2016
The Great Wall	2016
The Dark Tower	2017
Beauty and the Beast	2017
Fantastic Beasts: The Crimes of Grindelwald	2018
Aquaman	2018
Spider-Man: Far From Home	2019
Jumanji: The Next Level	2019
The Old Guard	2020
Bloodshot	2020
Love and Monsters	2020

a model $F : X \rightarrow [0, 1]^N$ that maps each trailer's feature vector to a probability distribution over the genres, such that $F(x) = [p(g1|x), p(g2|x), \ldots, p(gN|x)]$, where $p(g_i|x)$ is the probability of trailer x belonging to genre g_i. Data set D is a collection of labelled examples, where each example consists of a movie trailer and its corresponding genre label vector. Let's assume D contains m examples: $D = \{(x_1 y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ where $x_i \in X$ is a movie trailer represented by a feature vector of dimension d. Each trailer is a data point in the feature space X. $p_i \in {0, 1}^N$ is the genre probability vector for trailer x_i. Each element of p_i, p_{ij}, the probability that the trailer x_i belongs to genre g_j. The sum of all elements in p_i should equal 1 ($\sum_{j=1}^{N} p_{ij} = 1$), representing the probabilities of all possible genres for the trailer x_i.

A dataset of 25 movie trailers is prepared. Three annotators assigned genre probabilities to each trailer. They watched the trailers and assessed the likelihood of each trailer belonging to different genres by providing percentage probabilities. Annotators are given a list of possible genres and brief explanations to guide their judgment. The dataset thus includes movie trailers and their corresponding genre probability distributions, with the probabilities summing up to 1 (after conversion of percentages to probabilities). The annotation protocol can be found in Appendix A. The annotation quality is assessed using Fleiss' Kappa method as explained in Appendix B. Table 1 lists the selected movies for the dataset.

We developed a web-based system utilizing the Vue.js front-end and Flask back-end frameworks. The system's basic elements are shown in Figure 1. The data processing phase involves the generation of multi-label classification. A multi-modality method using visual and textual features has been used to classify trailers into genres. The subsequent sections provide the details:

A. VISUAL GENRE CLASSIFICATION

A video is composed of a series of consecutive, static images referred to as frames, which collectively form the visual content of the video. The video frames are classified into each genre for classification. We opted not to process every video frame to mitigate computational costs. Instead, we selected frames for processing by skipping every 10th frame throughout the trailer.

Each selected frame is then classified by assigning multiple labels and probabilities corresponding to each genre. A separate dataset was created for this classification by collecting images corresponding to Action, Adventure, Romance, Family, Fantasy, and Horror. The feature extraction process employs the pre-trained MobileNet model [35], excluding the top classification layer. MobileNet is designed for efficient and high-performance image classification and object detection on mobile and embedded devices. It achieves its efficiency by employing depth-wise separable convolutions, which reduce computational complexity and the number of parameters. Each labelled image within the images dataset undergoes pre-processing, involving resizing to fit the MobileNet input dimensions and scaling pixel values to the range of $[-1, 1]$. Subsequently, the pre-processed image is passed through the MobileNet model to capture features from a specific layer just before global average pooling. These extracted features are flattened to generate a feature vector for each image. The classifier network consists of fully connected layers. The input to this classifier is the flattened feature vector derived from MobileNet, and the output layer produces a binary vector signifying the presence or absence of each genre label. The dataset is partitioned into training (70%), validation (15%), and test (15%) sets. Throughout the training process, feature vectors traverse the classifier network, with binary cross-entropy loss applied to predict genre labels. The Adam optimizer is utilized, and validation loss and accuracy are monitored to mitigate overfitting. Data augmentation techniques such as rotation, flipping, and zooming are applied to augment the dataset and enhance generalization.

A similar pre-processing procedure is executed for every frame within a movie trailer, mirroring the steps taken with the training images. The pre-processed frame is subsequently passed through the MobileNet model to extract features. These extracted features are then forwarded through a trained classifier to generate genre predictions for that frame. The number of frames where each genre is predicted to be present throughout the trailer is counted. The average probability of each genre's presence is computed by dividing this count by the total number of frames that predicted the genre.

B. TEXTUAL GENRE CLASSIFICATION

Similar to visual features, a small textual dataset corresponding to each genre was collected. The textual features are
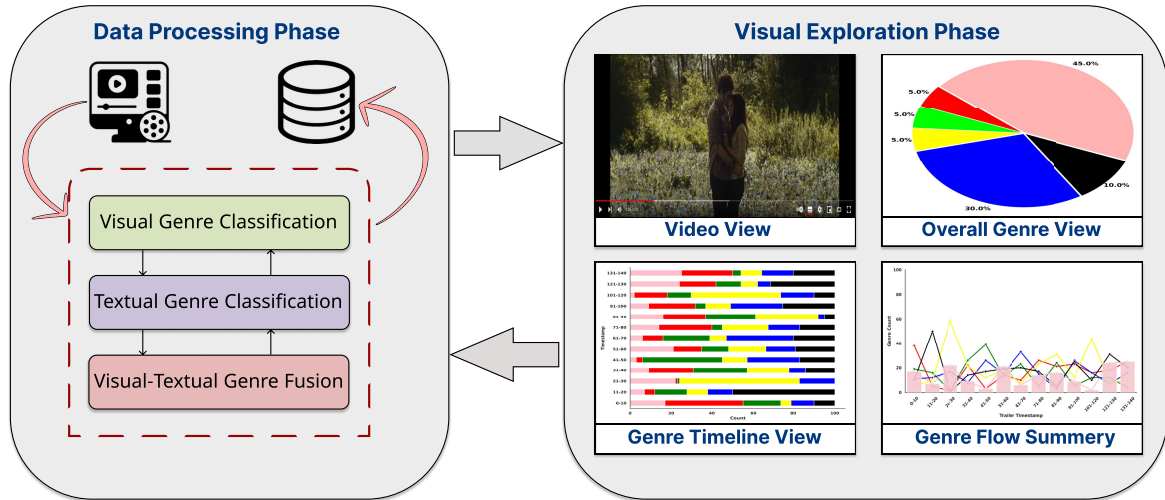
pre-processed by eliminating stop-words and punctuation and converting the text to lowercase. The processed text is then tokenized into segments to prepare it for BERT's [25] input format. BERT is known for its bidirectional context understanding, which pre-training on vast text data enables. It has performed brilliantly on diverse language tasks. A pre-trained BERT-base model derives contextualized embeddings for the textual segments. The tokenized segments are passed through the BERT model to obtain contextualized embeddings for each token. These embeddings captured the nuanced relationships and semantics present in the textual features. A classifier network is designed to predict genre labels based on BERT-derived textual feature representations. This network consisted of fully connected layers. The BERT-derived embeddings serve as input to the classifier, while the output layer comprises a binary vector indicating the presence or absence of each genre label. The dataset is split into training (70%), validation (15%), and test (15%) sets. The classifier is trained using the training set, with BERT-derived embeddings passed through the classifier network. Binary cross-entropy loss is utilized for each genre label's prediction, and the Adam optimizer facilitated weight updates. To ensure generalization, validation loss, and accuracy are monitored throughout training. For each frame in a movie trailer, the corresponding textual features are pre-processed, tokenized, and passed through the pre-trained BERT model to obtain contextualized embeddings. These embeddings are then fed into the trained textual classifier to predict genre labels for that frame. The counts of frames in which each genre label is predicted to be present across the entire trailer were tabulated. The average probability of each genre's presence was calculated by dividing the counts of frames with predicted genre labels by the total number of frames.

C. COMBINING PROBABILITIES FROM VISUAL AND TEXTUAL PREDICTIONS

The weighted average method is used to combine genre probabilities obtained from visual and textual classifiers.

Suppose w_v is the weight assigned to the predicted probability of the visual-based method, and w_t is the weight assigned to the predicted probability of the textual-based method. In that case, the combined probability $p_{combined}(g)$, for genre g is given by Eq. (1):

$$p_{combined}(g) = (w_v \cdot p_{visual}(g)) + (w_t \cdot p_{textual}(g)) \quad (1)$$

$p_{visual}(g)$ is the probability of the genre g based on visual features. $p_{textual}(g)$ is the probability of the genre g based on visual features. w_v and w_t are the weights of the visual and textual methods, respectively. The values of w_v and w_t should sum up to 1. We assigned an equal weight to w_v and w_t in our experiments.

IV. VISUALIZATION DESIGN

The visual exploration phase encompasses four primary perspectives: the video view, the overall genre view, the genre timeline view, and the genre flow summary view. In this section, the details of visualization design and all perspectives are provided. Figure 2 shows the main menu of the system. The user can browse a movie trailer and select one or more views to display.

Figure 3 shows the visualization made for the movie "The Twilight Saga: Eclipse" trailer when all views are selected. The video view (Figure 3a) displays the video we are examining. This view helps to provide evidence about the correction of the genre detection process as the users can directly watch the video along with the other visualizations related to genre detection. The overall genre view (Figure 3b) represents how different genres are distributed in the trailer, expressed as a percentage for each genre, thus providing an overview of the genre composition within the trailer, showing the proportion of content dedicated to each genre. The genre timeline view (Figure 3c) uses different colours to stack the genres on top of each other, representing the contribution of each genre to the overall content of the movie trailer during different time intervals. The genre flow view (Figure 3d) encapsulates the changes over time, focusing on one genre
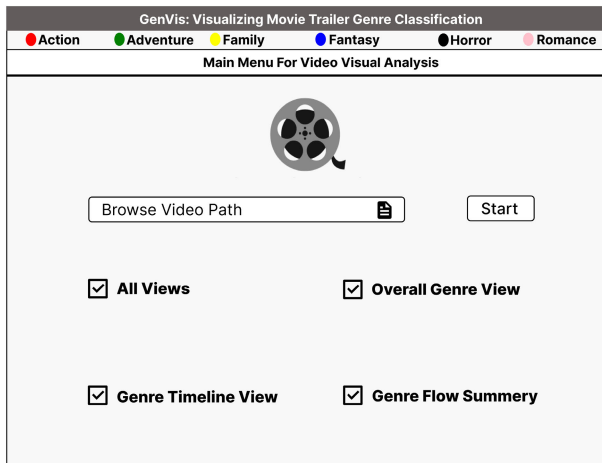
FIGURE 2. GenVis main menu.

within the chosen trailer. Figure 4 shows the scenario where the user only selected the overall genre view.

A. DESIGN RATIONALES

We adhered to the following three design principles in crafting our proposed system:

- **Intuitive Design and Encoding:** The system primarily targets film enthusiasts with limited data visualization experience. To meet the needs of this audience, the system employs straightforward visual designs that utilize familiar metaphors. This approach makes understanding the information regarding genres easier.
- **Seamless Interactions with Immediate Feedback:** The system has been engineered to provide smooth user interactions, ensuring an easy exploration experience. Users can focus on the highlighted content without any distractions.
- **Multi-Scale Visual Exploration:** Extracting genre information from videos yields a variety of content. The system enables multi-scale exploration, presenting an extensive overview and intricate details. This approach aids in swift and comprehensive comprehension of video content from a genre-specific perspective.

B. COLOUR SCHEME

For a unified colour encoding, the following colour scheme is used: red for Action, green for Adventure, pink for Romance, blue for Fantasy, yellow for Family, and black for Horror. The chosen colours for each genre enhance the emotional and thematic essence of the respective movie categories. Red, symbolizing energy and intensity, aligns with the Action genre's fast-paced and thrilling nature. Green, which stands for growth and exploration, matches the excitement of going on new adventures and exploring unknown places. Pink represents love and tenderness and captures the emotions relevant to romance films. The blue colour is often associated with mystery and the unknown and thus makes a good representation of fantasy movies' imaginative and fantastical

elements. Yellow signifies warmth and positivity and is usually associated with the joyful and uplifting atmosphere of family-oriented films. Black is often connected with darkness and fear, thus a suitable representation of horror movies.

C. VIDEO VIEW

The video view is designed to allow users to view movie trailers and other visualizations, which helps understand the genre identification process. In this view, users can adjust the video's speed and pause the video. Additionally, users can pause a video, effectively stopping it at a specific time. One unique feature of this "Video View" is the ability to analyze individual video frames. When a user pauses a video, the system provides visualization, giving users insights into the genre of the movie at the moment in the video when they pause it. In summary, the "Video View" tool allows users to watch movie trailers with advanced control over playback speed, pause and analyze individual frames, and access genre-related information when they pause the video. This functionality enhances the user's experience by providing more interactive and informative options while watching trailers.

D. OVERALL GENRE VIEW

The Overall Genre View is a visualization component that offers users a comprehensive analysis of movie genres detected within a trailer. This view presents the distribution of different genres, providing a quick and informative overview of the trailer's content. The user has the flexibility to view the distribution of genres as pie, lollipop, or polar charts as shown in Figure 5. In addition to selecting the chart type, users have the option to customize the number of genres they wish to view by adjusting a top genre slider located at the top of the interface. For instance, if the user chooses to display only a subset of the total six genres, such as four, the pie chart will prioritize showcasing the top four genres with the highest probabilities. Meanwhile, the remaining two genres will be visually de-emphasized or blurred within the pie chart (Figure 5 a). Similarly the bottom two genres will be blurred in Lollipop and Polar charts as well (Figure 5 b and c).

This view enhances user understanding, engagement, and decision-making. Since modern movies often blend elements from various genres, identifying a single genre can be challenging. The percentage view acknowledges this blending, informing users about the trailer's diverse aspects. The percentage distribution helps users form more informed expectations about the movie, preparing them for a multi-dimensional experience. This anticipation is valuable, especially since single-genre labels can be misleading when a movie defies conventional norms. For users interested in movie analysis, the percentage view offers insights into the trailer's genre composition, enabling more in-depth discussions and reviews. Presenting multiple genres with percentages caters to a broader audience and captures the interest of diverse viewers. The utilization of various types of charts ensures an engaging and user-friendly experience.

FIGURE 3. Visualization system for genre classification (sorted on "Action").

Additionally, the ability to customize the number of genres offers users enhanced flexibility and control over their data visualization experience. By allowing users to select the specific number of genres they want to view, they can focus on the most relevant information for their analysis or presentation.

E. GENRE TIMELINE VIEW

It is important to give users an overview of the genre's evolution. Therefore, a genre timeline view is designed to give users a summary of genre evolution in the trailer. Figure 3c illustrates that each horizontal stacked bar corresponds to the distribution of genres shown against a 10-second time interval. The users are provided with the flexibility to change the time interval. Moreover, the users can customize the genre types they want to focus on by selecting the check boxes on top of this view, giving them greater control over their analysis's specificity. For instance, Figure 6 shows the genre timeline view for different selections of genres.

The length of each area in the stacked bar chart indicates a particular genre's occurrence in that period. This layout can give users basic genre information about all the genres in the trailer and the proportions of each genre type. The name of the genre initially sorts the horizontal stacked bars. This view

helps users compare different types of genres more easily. By default, the horizontal stacked bars are initially arranged according to the order in the legend. However, if users choose to organize the horizontal stacked bars based on a particular genre, bars of that genre move to the left side for simple comparison.

F. GENRE FLOW SUMMARY

The genre flow summary shows the progression of genres over time. The users also have the option to select and visualize a complex genre flow (Figure 7) or simplified genre flow summary (Figure 8). The complex genre flow summary (as shown in Figure 3d) consists of a time interval, a genre band, and bar charts corresponding to the selected genre. A plot is created, incorporating a time interval, a genre band, and bar charts corresponding to the selected genre (Action genre is selected in Figure 3d). The bar chart helps to see the variation in percentages of the selected genres at various time intervals. The genre band helps the user visualize the progression of all genres over time. The time interval along the horizontal axis of the plot is variable. It segments the timeline into manageable units, allowing users to observe genre changes over those intervals. The genre band is a visual representation that spans the entire time
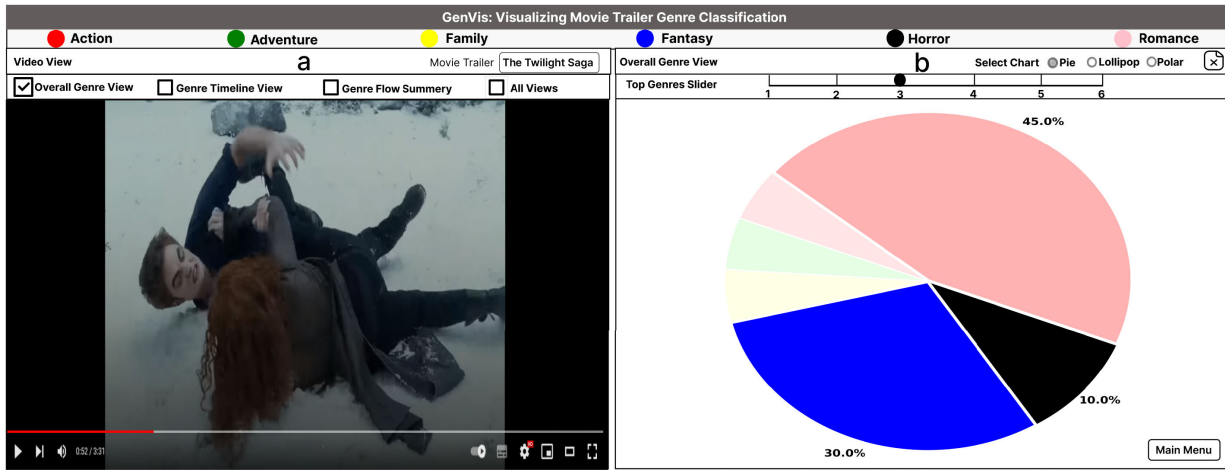
FIGURE 4. Visualization system with one selection of "Overall Genre View".
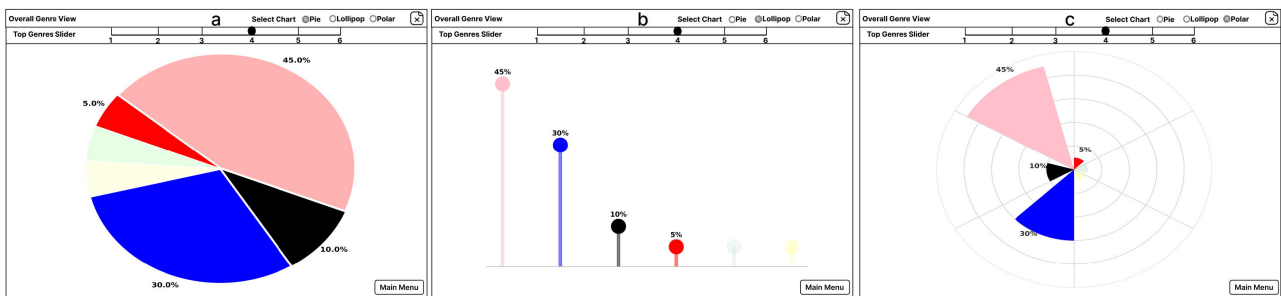

FIGURE 5. Overall genre view with different charts type and top genres slider selection.

axis. This band helps to visualize the relative prevalence of various genres at different times. For each time interval on the plot, a corresponding bar chart is positioned below the genre band. The heights of these bar charts represent the percentage of the selected genre within that time interval. The genre represented by each bar is colour-coded to match the corresponding genre in the genre band. The genre band gives the user a sense of how genres have evolved. As the user moves along the time axis, they can observe how the height of each genre segment changes, indicating how much screen time or attention that genre received during that period. By observing the shifts in the genre band and the bar charts, the user can identify trends, changes, and fluctuations of different genres over time. The simplified genre view (Figure 8) displays only the most prominent genre in each time interval, represented by bars, providing a clear and concise visualization of the dominant genres over time.

G. FRAME ANALYSIS VIEW

The users can pause a trailer and analyze the frame for genre detection. In this case, only the visual features are used to analyze the frame for genre detection. Only the video and the overall genre views are available for this type of analysis. The overall genre views show the distribution of genres only based on the selected frame. Figure 9 shows an example of this type

of analysis. The dominant genre in this particular frame is Horror, as seen in the Overall genre view. Users can choose the number of genres they prefer to view. Moreover, the users can select between pie and donut charts.

H. GENRE DETECTION ON MULTIPLE TRAILERS

The system allows the visualization of the distribution of genres across multiple trailers. For this, the users can select five videos for genre detection. Depending on the user's choices, the system employs a flexible visualization approach, presenting information through informative stacked bar graphs or intuitive pie charts. Figure 10 shows an example of an interaction for this type of visualization where the genre distribution of five movies is shown.

I. SORTING BY A PARTICULAR GENRE

The users can visualize the genre timeline view and genre flow summary by sorting a particular genre. In Figure 3, the genre timeline view was sorted on Action, and thus, the bars of the stacked bar chart start from Action. Similarly, in the Genre Flow view, the focus is on genre Action. In Figure 11, in the genre timeline view, the users selected to view time progression based on all genres except Action with a time interval of 10 seconds sorted on Romance. So, the bars of the stacked bar chart start with Romance. Finally, it is to be noted
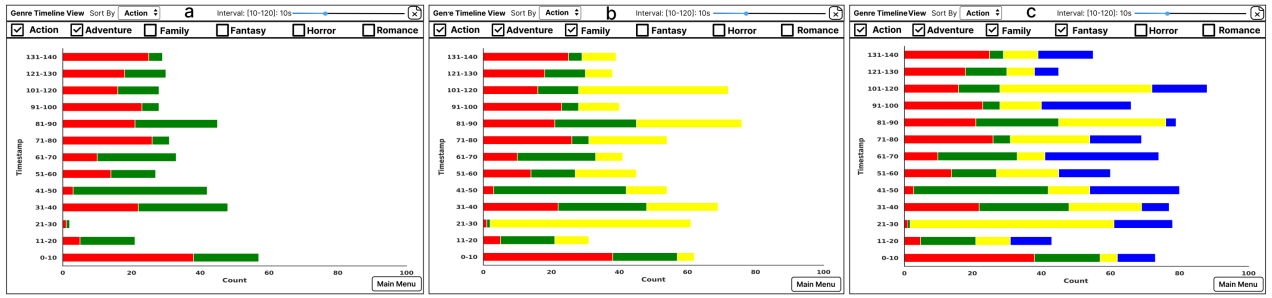
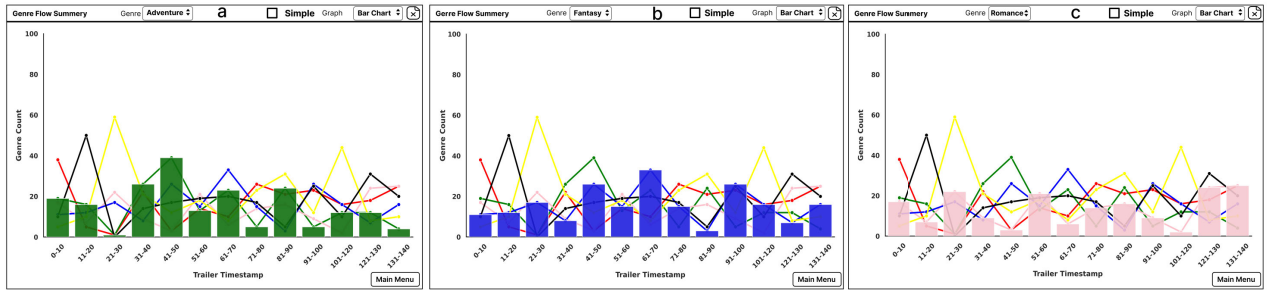**FIGURE 6.** Genre timeline view with different genres selection option.

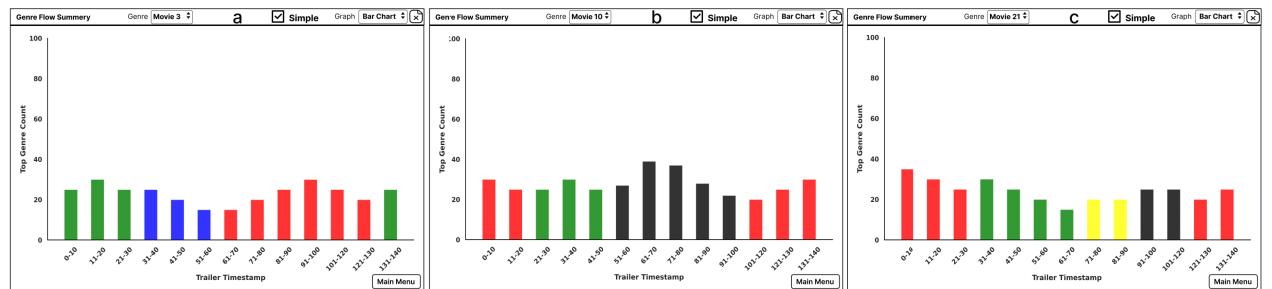**FIGURE 7.** Genre flow summery with one genre selection in timeline "Complex View".

FIGURE 8. Genre flow summery with top genre in each timeline for different movies "Simple View".

that the sorting option is not available for simple genre flow summaries, and the option to sort gets disabled if the user prefers to view the simple view.

V. EXPERIMENTS AND RESULTS

A. EVALUATION OF VIDEO GENRES CLASSIFICATION

Three methods have been used to evaluate multi-label genre classification. The first and second methods calculate the mean squared error and cross-entropy loss between human and machine annotation for each movie. The third method measures each genre's precision, recall, and F1 score. Moreover, a user study has also been conducted to determine the efficacy of visualization

1) MEAN SQUARE ERROR BETWEEN HUMAN AND MACHINE ANNOTATIONS

For each genre j, the mean squared error (MSE) measures the average squared difference between that genre's actual and predicted probabilities across all movies. This shows how

well the machine's predictions match the ground truth for each genre. Let N be the total number of movies and K be the total number of genres. For each genre j, the MSE for each movie i is calculated as follows by Eq. (2):

$$MSE_i = \frac{1}{N} \sum_{j=1}^{N} (TrueProb_{ij} - PredProb_{ij})^2 \qquad (2)$$

where N is the number of genres (in this case, 6) for each movie. $TrueProb_{ij}$ is the true probability of genre j for movie i. $PredProb_{ij}$ is the predicted probability of genre j for movie i. MSE_i is the Mean Squared Error for movie i, which represents the average squared difference between the true and predicted probabilities across all genres for that movie. Average MSE is given by Eq. (3):

$$AverageMSE = \frac{1}{M} \sum_{i=1}^{M} MSE_i \qquad (3)$$

where M is the total number of movies. MSE_i is the Mean Squared Error for movie i, as calculated earlier. The

FIGURE 9. Frame level analysis.

minimum possible value of MSE can be 0 when human and machine-generated annotations assign the exact probabilities to each genre. The maximum value of MSE can be 1 when the human and machine-generated annotated probabilities are as dissimilar as possible. Thus, an MSE value closer to 0 reflects a good correlation between human annotations and machine-generated probabilities. On the other hand, a value closer to 1 means a big difference in human and machine annotations. The average MSE value between human and machine annotations in our experiments is 0.019. A value of 0.019 indicates a relatively small discrepancy between the two sets of annotations, implying that the machine's performance is aligned with human judgments. The MSE values for all individual movie trailers are depicted in Figure 12.

2) CROSS ENTROPY LOSS BETWEEN HUMAN AND MACHINE ANNOTATIONS

Cross-entropy loss has been used as a second metric to measure the difference between human and machine annotations. Cross-entropy loss is a measure that quantifies the dissimilarity between the two sets of annotations. The Cross-Entropy Loss for a single movie i can be calculated as Eq. (4):

$$CE_i = -\sum_{j=1}^{N} \left(\text{TrueProb}_{ij} \cdot \log(\text{PredProb}_{ij}) + (1 - \text{TrueProb}_{ij}) \cdot \log(1 - \text{PredProb}_{ij}) \right) \quad (4)$$

where N is the total number of genres, $TrueProb_{ij}$ is the true probability of genre j for movie i, $PredProb_{ij}$ is the predicted probability of genre j for movie i. The average Cross-Entropy Loss for all movies is given by Eq. (5):

$$\text{Average CE} = \frac{1}{M} \sum_{i=1}^{M} CE_i \quad (5)$$

The average cross-entropy values range between 0 and ∞, where a value close to 0 is considered a good value that shows the closeness of human and machine-generated annotations. The average cross-entropy loss value in our experiments is
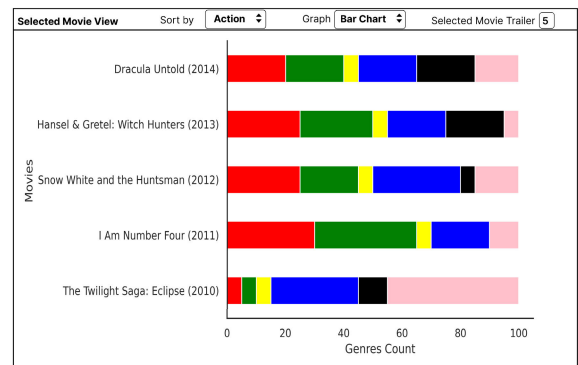

FIGURE 10. Multi-label Genre Visualization for multiple Movie Trailers.

1.32. Figure 13 shows the cross-entropy loss values for each movie. Lower cross-entropy loss values generally indicate better model performance, suggesting that the predicted probabilities are closer to the true labels.

When the mean square error or cross-entropy loss is high, it signifies a substantial disparity between machine-generated and human annotations. As an example, in the case of movie 25, the cross-entropy loss value is exceptionally high. The average human annotation for movie 25 (Movie: Love and Monsters) is Action = 13.33%, Adventure = 31.67%, Romance = 31.67%, Fantasy = 10% Family = 3.33% Horror = 3.33%, whereas the machine annotation is Action = 10%, Adventure = 40%, Romance = 20%, Fantasy = 0% Family = 20% Horror = 10%. The human and machine annotations for the predicted genre probabilities of this movie exhibit noticeable disparities. These disparities highlight the complexity of the task and indicate that humans often consider context, themes, and subtle nuances while machines rely on patterns and data.

3) PRECION, RECALL, AND F1 SCORE FOR EACH GENRE

Precision, recall, and F1-score metrics are calculated for each genre in multi-label genre classification. For a given genre g, precision P is the ratio of the number of correct positive predictions for that genre to the total number of positive predictions made. The formula for precision
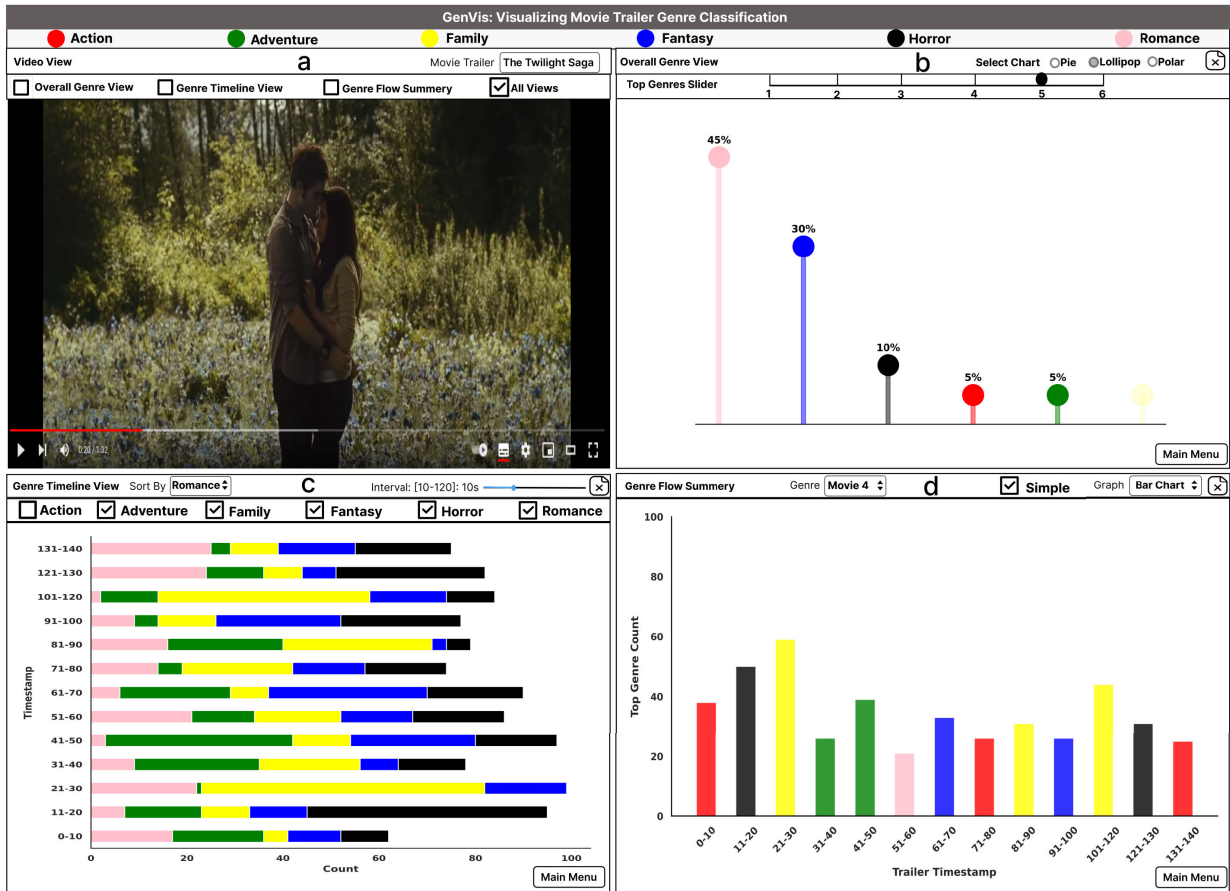
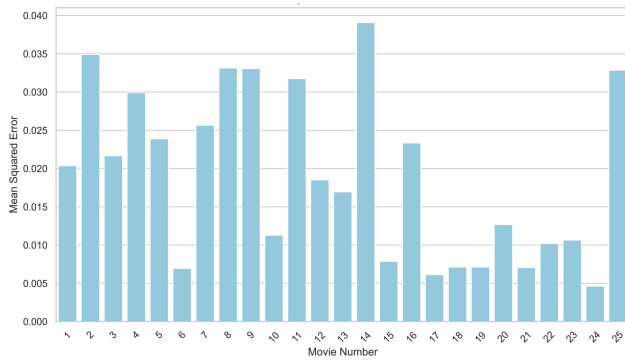**FIGURE 11.** Visualization of genres sorted on romance.

FIGURE 12. Mean Squared Error between Human and Machine Annotation for each Movie.

FIGURE 13. Cross Entropy Loss between Human and Machine Annotation for each Movie.

is given by Eq. (6):

$$P_g = \frac{TP_g}{TP_g + FP_g} \quad (6)$$

For a given genre g, recall (R) is the ratio of the number of correct positive predictions for that genre to the total number of actual instances of that genre. It can be written as Eq. (7):

$$R_g = \frac{TP_g}{TP_g + FN_g} \quad (7)$$

where:

- TP_g (True Positives for genre g is the number of times the genre g was correctly predicted.
- FP_g (False Positives for genre g is the number of times the genre g was incorrectly predicted when it was not the actual genre.
- FN_g (False Negatives for genre g) is the number of times the genre g was not predicted when it should have been.

The F1 score ($F1$) for a given genre g combines precision and recall into a single metric using the harmonic mean of precision and recall for that genre. It is given by Eq. (8):

$$F1_g = 2 \times \frac{P_g \times R_g}{P_g + R_g} \tag{8}$$

Table 2 presents the performance metrics for genre classification across different top selections (from Top 1 to Top 5) for each movie genre. In Table 2, "Top 1" denotes the scenario where only the most probable genre is selected for each movie. "Top 2" extends this to include the two highest probability genres. Similarly, "Top 3", "Top 4", and "Top 5" progressively consider the three to five most probable genres for each movie. Each "Top" level reflects a broader selection of potential genres based on descending probability.

It can be seen that the Action, Adventure, and Fantasy genres consistently show high scores across all metrics, especially as the number of top genres considered increases. For instance, Action and Fantasy reached a precision and recall 1.00 in the Top 5 evaluation, leading to perfect F1 scores. This suggests the classifier can correctly predict these genres when more genre labels are allowed per movie. The genre Romance shows a notable improvement in recall from Top 4 (1.00) to Top 5 (0.929), albeit with a slight drop in precision. The Family genre significantly improves all metrics as the number of top genres considered increases, moving from no correct predictions in the Top 1 to much higher scores in the Top 5 (precision of 0.905, recall of 0.864).

The genre Horror exhibits the lowest scores, with a modest peak in the Top 5 (precision of 0.643, recall of 0.643). This is because the human annotators did not assign Horror as a Top 1, Top 2, and Top 3 label, which explains why precision, recall, and F1-score are zero in the Top1, Top2, and Top3 genre evaluations. Similarly, in the Top 4, the human annotators only assigned the horror label to 4 movies, which is the reason for the low scores. Similarly, the genre Family appears 0, 2, and 6 times in Top 1, Top 2, and Top 3 human annotations, respectively, thus explaining the low scores. However, as the allowed number of genre labels in human and machine annotations increases, the classifier's predictions improve. The overall trend shows that precision, recall, and F1 scores generally increase or stabilize as more genre labels per movie are considered.

4) GENRE WISE ANALYSIS

In multi-label classification, where each instance can belong to multiple genres, traditional confusion matrices do not directly apply in the same way they do in multi-class settings. Therefore, we computed the label co-occurrence matrices for human annotations and machine predictions to analyze genre confusion. The co-occurrence matrix displays the frequency of pairs of movie genres appearing together in a dataset. For simplicity, we only provide the label co-occurrence matrix for the top 3 genres. The co-occurrence matrix for the top 3 genres can be seen in Table 3. The format used in each entry of the table is (Human, Machine) for each genre pairing. The

co-occurrence matrix shows how often each pair of genres is annotated or predicted together.

Table 3 shows that the Action and Adventure genres have strong co-occurrences with each other (18, 14). Fantasy frequently co-occurs with both Action and Adventure genres in human and machine predictions (Action and Fantasy: 15, 11; Adventure and Fantasy: 16, 16). This trend suggests that Fantasy elements are often integrated with Adventure and Action genres, possibly due to shared thematic elements like heroism and extraordinary settings. Romance is frequently confused with Family and Adventure, while Fantasy is often mistaken for Adventure and Horror. Action is often missed when predicting Romance and other genres like Family and Adventure. Finally, the zeros in the Horror row of the human annotations co-occurrence matrix suggest that the genre Horror was not annotated for any movie by humans in your dataset.

B. EVALUATION OF VISUALIZATION

User studies have been widely used for evaluating visualization quality [32], [36] and subjective video analysis systems [37], [38]. To conduct the user study, 43 undergraduate students majoring in computer science were asked to assess the effectiveness of the proposed visualization system by filling out a questionnaire. The average age of the participants is 22.28, with 25 male and 18 female participants. All participants had sufficient exposure to watching movies.

The questionnaire and the user responses are available on [39]. The responses were then analyzed to evaluate the system's usability, visual appeal, impact on genre understanding, and potential for future use. The average results are shown in Table 4.

The evaluation of the system reveals notable insights into its usability and effectiveness across various dimensions. Users generally found the system comprehensible, as evidenced by the satisfactory score of 3.83 for system understanding. Moreover, the system's visual appeal received positive feedback, with a rating of 3.91. Navigation within the system was perceived as intuitive, garnering a favourable score of 3.95. Notably, the genre view was highly comprehensible to users, achieving a score well above average at 4.27. While users demonstrated a strong understanding of pie plot representations, scoring 3.94, comprehension of polar plots appeared comparatively lower at 3.60. However, the lollipop plot within the genre view was well-received, scoring 4.19. Genre number selection in genre and timeline views received moderately positive ratings, suggesting room for improvement. The significance of the genre flow summary was generally acknowledged, with a score of 3.90. Furthermore, detailed and simplified views within the genre flow summary were well-understood, scoring 4.20 and 4.04, respectively. Overall, the system garnered positive feedback, culminating in an impressive overall rating of 4.25, indicating its efficacy in facilitating user interaction and comprehension.

TABLE 2. Recall, Precision, F1 score for top 1 to top 5 genre classification.

Genre	Top 1			Top 2			Top 3			Top 4			Top 5		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Action	0.875	0.500	0.636	0.909	0.714	0.800	1.000	0.667	0.800	0.941	0.667	0.780	1.000	0.960	0.980
Romance	1.000	0.500	0.667	0.667	0.500	0.571	0.833	0.625	0.714	0.667	1.000	0.800	0.722	0.929	0.813
Family	0.000	0.000	0.000	0.167	0.500	0.250	0.375	0.500	0.429	0.545	0.429	0.480	0.905	0.864	0.884
Adventure	0.000	0.000	0.000	0.800	0.632	0.706	0.857	0.857	0.857	0.955	0.875	0.913	1.000	0.960	0.980
Fantasy	0.286	0.500	0.364	0.455	0.455	0.455	0.750	0.789	0.769	0.958	0.958	0.958	1.000	0.960	0.980
Horror	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.182	0.500	0.267	0.643	0.643	0.643

TABLE 3. Combined co-occurrence matrix of human annotations and machine predictions (Top 3 Genres). The format used is (Human, Machine) for each genre pairing.

	Action	Romance	Family	Adventure	Fantasy	Horror
Action	(21, 14)	(6, 1)	(3, 1)	(18, 14)	(15, 11)	(0, 1)
Romance	(6, 1)	(8, 6)	(2, 3)	(4, 2)	(4, 4)	(0, 2)
Family	(3, 1)	(2, 3)	(6, 8)	(4, 6)	(3, 5)	(0, 1)
Adventure	(18, 14)	(4, 2)	(4, 6)	(21, 21)	(16, 16)	(0, 4)
Fantasy	(15, 11)	(4, 4)	(3, 5)	(16, 16)	(19, 20)	(0, 4)
Horror	(0, 1)	(0, 2)	(0, 1)	(0, 4)	(0, 4)	(0, 6)

TABLE 4. Summary of user survey results on visualization.

Criteria	Score (1-5)
Understanding of system	3.83
Visual Appeal	3.91
Navigation Ease	3.95
Overall Genre View Understanding	4.27
Overall Genre View: Pie Plot	3.94
Overall Genre View: Polar Plot	3.60
Overall Genre View: Lollipop Plot	4.19
Overall Genre View: Genre Number Selection	3.77
Timeline View Understanding	3.80
Timeline View: Genre Number Selection	4.05
Genre Flow Summary Significance	3.90
Genre Flow Summary: Detailed View	4.20
Genre Flow Summary: Simplified View	4.04
Overall Rating	4.25

The user study results suggest that the participants have well-received the developed visualization system for genre detection in movie trailers. The system's visualizations successfully communicated genre distribution and temporal patterns, enhancing the audience's comprehension of movie genres. The positive scores for visual appeal, usability, and likelihood of future use indicate that the system has the potential to be a valuable tool for users seeking to explore and understand movie genres.

C. LIMITATIONS

Since the study focused on movie trailers rather than full-length films, it may have limited ability to capture the full scope and complexity of genres. Trailers, being brief, may not encompass all the nuances of genres present in complete movies. Consequently, the findings and visualizations from trailers might not fully represent the genre dynamics of complete films. Furthermore, the visualization mechanism developed for trailers may not be directly transferable to full-length movies due to differences in content duration, complexity, and viewer engagement. The condensed nature of trailers allows for specific visualization features that might not be feasible or effective in full-length films. Thus, the current visualization system may not be suitable for analyzing the genre dynamics of complete movies.

Many modern movie genre detection techniques leverage multi-modal data sources, including posters, audio content, and trailers. However, our study primarily relied on subtitles of trailers and visual cues, excluding the use of poster images and audio content. This limitation may affect the comprehensiveness and accuracy of genre detection compared to methods that incorporate a wider range of data modalities. While our approach has its strengths, such as focusing on subtitles and visual clues, it is important to acknowledge that the exclusion of poster images and audio data may limit the overall effectiveness of our genre detection model.

Our current research focused exclusively on six movie genres: Action, Adventure, Romance, Family, Fantasy, and Horror. While these genres are widely recognized and prevalent, it is important to acknowledge that cinema encompasses various genres and sub-genres. The omission of other genres, such as drama, science fiction, comedy, and documentary, among many others, means that our genre detection system may not capture the full diversity of movie genres. This limitation implies that our findings and visualizations are specific to the selected genres and may not apply to movies outside this limited scope.

VI. CONCLUSION

The GenVis visualization system addresses the complexity of multi-label movie genres extracted from trailers. GenVis visualizes multi-label genre results, providing insights into each trailer's genre dynamics. It offers four main views: the video view for direct trailer engagement, the overall genre view displaying genre percentage distribution, the genre timeline view summarizing genre evolution, and the genre flow summary for detailed temporal analysis of specific

genres. Users can pause on frames for deeper analysis, sort results by genre, and concurrently process genre distributions for up to five videos. The system ensures accurate genre classification through text and visual features. The video view enables comprehensive trailer exploration, enhancing genre comprehension. Transparent visualization of multiple genre labels with probabilities supports informed assessment. The genre timeline view illuminates genre evolution while sorting capabilities streamline analysis. Concurrent processing enhances cross-trailer comparison. Evaluation, including mean squared error and cross-entropy loss, confirms the system's multi-label genre detection proficiency. A user study validates GenVis's effective impact on presenting complex genre distribution patterns and temporal changes. These outcomes underscore GenVis's success in categorizing genres within trailers and offering valuable analytical insights.

Looking ahead, our research aims to expand by incorporating additional video genres. This extension will deepen insights into how visual and auditory elements differ across genres. We also plan to explore visualization methods in non-movie video contexts like social media and consumer-generated content. While similar methodologies can be adapted, adapting to dynamic videos might introduce new genres. Further, we can perform more detailed analyses, including shot and scene understanding, event recognition, and emotion recognition, enhancing our genre-centric approach.

APPENDIX A
ANNOTATION PROTOCOL

The objective of this annotation protocol is to collect genre probability annotations for a set of movie trailers. Annotators watch each movie trailer and assign percentages to each possible genre based on their judgment and understanding of the content.

1) The dataset of 25 movie trailers used for annotation is introduced, highlighting its diversity and covering various movie genres and themes.
2) Annotators are presented with a list of possible genres, and each genre is clearly defined with its characteristics and examples to facilitate understanding. The list is comprehensive and covers all relevant genres for the movie trailers.
3) The importance of paying close attention to the content of each trailer is emphasized. Annotators must focus on visual, audio, and textual cues to assess genre probabilities. They assign percentages to each genre, representing the likelihood of occurrence.
4) Annotators are instructed to watch each movie trailer thoroughly, considering various aspects such as visuals, audio, narrative, and themes. For each trailer, they assign a probability percentage to each genre based on how likely they think the trailer belongs to that genre. Emphasis is placed on ensuring that the probabilities

sum to 100% to represent the certainty that the trailer belongs to one of the listed genres.
5) The annotation protocol acknowledges that some trailers might exhibit characteristics of multiple genres, leading to ambiguity. In such cases, annotators are instructed to assign probabilities to all relevant genres, reflecting the uncertainty.
6) A clear scale for assigning probabilities (e.g., 0 to 100) is established, and the meaning of low, medium, and high probabilities is explained.
7) Structured annotation forms are provided for each movie trailer, with fields for annotators to input the probabilities for each genre.
8) Annotators are advised to avoid discussing their annotations with others to maintain independence and avoid bias.

APPENDIX B
EVALUATING QUALITY OF ANNOTATION

For evaluating the quality of the annotations in the annotation protocol, an Inter-annotator agreement (IAA) measure has been used. IAA measures the level of agreement or consistency between different annotators when assigning annotations to the same data. It provides insights into the reliability and quality of the annotations. Fleiss' Kappa is suitable for situations where there are multiple annotators (more than two). The following steps were used for evaluating the inter-annotator agreement in the annotation of trailer videos using Fleiss' Kappa method:

1) A dataset of 25 movie trailers is prepared, including genre probability annotations from three different annotators (A, B, and C). Thus, each movie trailer is associated with a vector of genre probability values assigned by each annotator.
2) An annotation matrix of dimension dimensions (nxm) shows the annotations of the three annotators for each movie trailer where n represents the count of movie trailers and m represents the number of possible genres. Each matrix element (i, j) represents the count of annotators assigning genre j to movie trailer i with probability values.
3) The value of Observed agreement A is calculated as: $A = \frac{\sum_i n_i}{N \cdot k}$, where n_i denotes the number of annotators agreeing on genre i for a movie trailer, N denotes for the total number of movie trailers, and k denotes the number of annotators ($k = 3$ in this case). The value of A is calculated for each movie trailer, which shows the proportion of agreements among annotators for that particular trailer.
4) The calculation of agreement expected by chance (E) relies on the probabilities of each genre being assigned by each annotator. The formula for agreement expected by chance is $E = \frac{\sum_j p_j^2}{N \cdot k}$, where p_j signifies the proportion of genre j being assigned by all annotators.

5) Fleiss' Kappa (κ) is computed as the normalized difference between observed agreement and agreement expected by chance. The formula for Fleiss' Kappa is $\kappa = \frac{A-E}{1-E}$.

6) Fleiss' Kappa value spans from -1 to 1, where 1 indicates perfect agreement, 0 indicates agreement by chance, and negative values suggest less agreement than expected by chance.

7) The interpretation of Fleiss' Kappa value is pivotal in evaluating the level of agreement among annotators. Elevated positive values (approximating 1) signal enhanced agreement, whereas values nearing 0 or negative values indicate limited agreement.

Through the application of Fleiss' Kappa method, the inter-annotator agreement in the genre probability annotations for the movie trailers can be quantitatively assessed. A high Fleiss' Kappa value would suggest notable consensus among the annotators, instilling confidence in the dataset's dependability and the aptness of the annotations for both training and assessing the genre extraction model. After some iteration, the Fleiss' Kappa value between the three annotators reached 0.21, indicating fair agreement.

DECLARATION OF COMPETING INTEREST
We do not have any conflict of interest.

ACKNOWLEDGMENT

REFERENCES
[1] A. S. Mehal, K. Meena, R. B. Singh, and P. G. Shambharkar, "Movie genres and beyond: An analytical survey of classification techniques," in *Proc. 5th Int. Conf. Trends Electron. Informat. (ICOEI)*, Jun. 2021, pp. 1193–1198.

[2] P. G. Shambharkar, A. Anand, and A. Kumar, "A survey paper on movie trailer genre detection," in *Proc. Int. Conf. Comput. Data Sci. (CDS)*, Aug. 2020, pp. 238–244.

[3] K. Kundalia, Y. Patel, and M. Shah, "Multi-label movie genre detection from a movie poster using knowledge transfer learning," *Augmented Human Res.*, vol. 5, no. 1, pp. 1–9, Dec. 2020.

[4] Kaggle. (2023). *Genre Classification Dataset IMDb*. Accessed: Aug. 19, 2023. [Online]. Available: https://www.kaggle.com/datasets/hijest/genre-classification-dataset-imdb

[5] Kaggle. (2023). *Movie Genre+Plot+Poster*. Accessed: Aug. 20, 2023. [Online]. Available: https://www.kaggle.com/datasets/aakashsaroop/movie-genreplotposter

[6] PLOS. (2023). *Genre Classification Using Deep Learning for LMTD Movies Dataset*. Accessed: Aug. 20, 2023. [Online]. Available: https://plos.figshare.com/articles/dataset/Genre_classification_using_Deep_Learning_for_LMTD_movies_dataset_/7759151/1

[7] Y. Deldjoo, M. G. Constantin, B. Ionescu, M. Schedl, and P. Cremonesi, "MMTF-14K: A multifaceted movie trailer feature dataset for recommendation and retrieval," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018, pp. 450–455.

[8] F. Z. Unal, M. S. Guzel, E. Bostanci, K. Acici, and T. Asuroglu, "Multilabel genre prediction using deep-learning frameworks," *Appl. Sci.*, vol. 13, no. 15, p. 8665, Jul. 2023.

[9] R. Montalvo-Lezama, B. Montalvo-Lezama, and G. Fuentes-Pineda, "Improving transfer learning for movie trailer genre classification using a dual image and video transformer," *Inf. Process. Manage.*, vol. 60, no. 3, May 2023, Art. no. 103343.

[10] Y. Yu, Z. Lu, Y. Li, and D. Liu, "ASTS: Attention based spatio-temporal sequential framework for movie trailer genre classification," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 9749–9764, Mar. 2021.

[11] A. Yadav and D. K. Vishwakarma, "A unified framework of deep networks for genre classification using movie trailer," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106624.

[12] G. S. Simoes, J. Wehrmann, R. C. Barros, and D. D. Ruiz, "Movie genre classification with convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 259–266.

[13] J. Wehrmann and R. C. Barros, "Movie genre classification: A multi-label approach based on convolutions through time," *Appl. Soft Comput.*, vol. 61, pp. 973–982, Dec. 2017.

[14] Y.-F. Huang and S.-H. Wang, "Movie genre classification using SVM with audio and video features," in *Proc. 8th Int. Conf. Act. Media Technol. (AMT)*, Macau, China. Berlin, Germany: Springer, Dec. 2012, pp. 1–10.

[15] H. Zhou, T. Hermans, A. V. Karandikar, and J. M. Rehg, "Movie genre classification via scene categorization," in *Proc. 18th ACM Int. Conf. Multimedia*, Oct. 2010, pp. 747–750.

[16] J. A. Wi, S. Jang, and Y. Kim, "Poster-based multiple movie genre classification using inter-channel features," *IEEE Access*, vol. 8, pp. 66615–66624, 2020.

[17] G. Barney and K. Kaya, "Predicting genre from movie posters," in *Proc. Stanford CS 229, Mach. Learn.*, 2019, pp. 1–6.

[18] N. K. Rajput and B. A. Grover, "A multi-label movie genre classification scheme based on the movie's subtitles," *Multimedia Tools Appl.*, vol. 81, no. 22, pp. 32469–32490, Sep. 2022.

[19] G. Portolese, M. A. Domingues, and V. D. Feltrim, "Exploring textual features for multi-label classification of Portuguese film synopses," in *Proc. 19th EPIA Conf. Artif. Intell.*, Vila Real, Portugal. Cham, Switzerland: Springer, Sep. 2019, pp. 669–681.

[20] A. M. Ertugrul and P. Karagoz, "Movie genre classification from plot summaries using bidirectional LSTM," in *Proc. IEEE 12th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2018, pp. 248–251.

[21] Z. Cai, H. Ding, J. Wu, Y. Xi, X. Wu, and X. Cui, "Multi-label movie genre classification based on multimodal fusion," *Multimedia Tools Appl.*, vol. 83, no. 12, pp. 36823–36840, Jul. 2023.

[22] R. B. Mangolin, R. M. Pereira, A. S. Britto, C. N. Silla, V. D. Feltrim, D. Bertolini, and Y. M. G. Costa, "A multimodal approach for multi-label movie genre classification," *Multimedia Tools Appl.*, vol. 81, no. 14, pp. 19071–19096, Jun. 2022.

[23] T. Bi, D. Jarnikov, and J. Lukkien, "Shot-based hybrid fusion for movie genre classification," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2022, pp. 257–269.

[24] L. Braz, V. Teixeira, H. Pedrini, and Z. Dias, "Image-text integration using a multimodal fusion network module for movie genre classification," in *Proc. 11th Int. Conf. Pattern Recognit. Syst. (ICPRS)*, Mar. 2021, pp. 200–205.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[26] J. Arevalo, T. Solorio, M. Montes-y-Gómez, and F. A. González, "Gated multimodal units for information fusion," 2017, *arXiv:1702.01992*.

[27] N. Ejaz, I. Mehmood, J. J. Lee, S. M. Ji, M. H. Lee, S. M. Anh, and S. W. Baik, "Interactive 3D visualization of soical network data using cloud computing," in *Proc. Int. Conf. Cloud Comput. Social Netw. (ICCCSN)*, Apr. 2012, pp. 1–4.

[28] G. Daniel and M. Chen, "Video visualization," in *Proc. 14th IEEE Visualization*, vol. 54. Washington, DC, USA: IEEE Computer Society, Oct. 2003, pp. 409–416, doi: 10.1109/VISUAL.2003.1250401.

[29] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim, "Bring it to the pitch: Combining video and movement data to enhance team sport analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 13–22, Jan. 2018.

[30] J. Wang, J. Wu, A. Cao, Z. Zhou, H. Zhang, and Y. Wu, "Tac-miner: Visual tactic mining for multiple table tennis matches," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 6, pp. 2770–2782, Jun. 2021.

[31] T. Lin, Z. Chen, Y. Yang, D. Chiappalupi, J. Beyer, and H. Pfister, "The quest for : Embedded visualization for augmenting basketball game viewing experiences," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 1, pp. 962–971, Jan. 2023.

[32] H. Zeng, X. Shu, Y. Wang, Y. Wang, L. Zhang, T.-C. Pong, and H. Qu, "EmotionCues: Emotion-oriented visual summarization of classroom videos," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 7, pp. 3168–3181, Jul. 2021.

[33] N. Mumtaz, N. Ejaz, S. Aladhadh, S. Habib, and M. Y. Lee, "Deep multi-scale features fusion for effective violence detection and control charts visualization," *Sensors*, vol. 22, no. 23, p. 9383, Dec. 2022.

[34] G. Chen, J. Huang, C. Wei, J. Yang, M. M. Scully, A. Sergeev, M.-C. Chen, S. G. Krantz, P. Yao, T. Guo, and J. Wang, "Animal shapes, modal analysis, and visualization of motion (I): Horse and camel," *J. Geometric Anal.*, vol. 33, no. 10, p. 328, Oct. 2023.

[35] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[36] N. Lettieri, A. Guarino, D. Malandrino, and R. Zaccagnino, "The sight of justice. Visual knowledge mining, legal data and computational crime analysis," in *Proc. 25th Int. Conf. Inf. Visualisation (IV)*, Jul. 2021, pp. 267–272.

[37] U. A. Khan, N. Ejaz, M. A. Martínez-del-Amor, and H. Sparenberg, "Movies tags extraction using deep learning," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2017, pp. 1–6.

[38] N. Ejaz, U. A. Khan, M. Á. M. del Amor, and H. Sparenberg, "Deep learning based beat event detection in action movie franchises," in *Proc. 10th Int. Conf. Mach. Vis. (ICMV)*, Apr. 2018, pp. 35–42.

[39] F. Shaukat. (2024). *Visualizing Genre Detection*. Accessed: May 16, 2024. [Online]. Available: https://github.com/faheemshaukat/Visualizing-Genre-Detection

ZEESHAN ASHRAF (Senior Member, IEEE) received the M.S.C.S. degree in computer networking and security services from UMT, Lahore, Pakistan, and the Ph.D. degree in computer science from IQRA University, Islamabad Campus, Islamabad, Pakistan.

He has more than 15 years of experience in the IT field. He is currently a CISCO CCNP (routing and switching) certified. He is the author of several books published by international publishers. His several research articles have been published in well-reputed international conferences and Q1 & Q2 journals having good impact factors. He is also an Associate Professor with the Computer Science Department, The University of Chenab, Gujrat, Pakistan. His main research interests include next-generation virtualized internet architecture, IPv6 routing, performance modeling, the IoT, optimization techniques, and security services in different networks.

FAHEEM SHAUKAT received the M.S. degree in computer science with a focus on artificial intelligence and software engineering from IQRA University, Islamabad Campus, Islamabad, Pakistan, where he is currently pursuing the Ph.D. degree in computer science.

With over five years of experience in machine learning and data science, he holds several certifications in these areas and has authored and published papers at esteemed conferences. He is also a Permanent Faculty Member with the Computer Science Department, IQRA University. His primary research interests include machine learning models, data science, and digital image processing.

MRIM M. ALNFIAI is currently an Associate Professor in information technology with Taif University, Saudi Arabia. Her research interests include assistive technology, human–computer interaction, accessibility, usable security, AI, and machine learning. She has published several papers at ISI journals and assistive technology, HCI, and accessibility conferences, including ASSETS, ANT, FNC, CIST, JAIHC, and ICCA. Her current research interest includes designing accessible tools for visually impaired people, including people with no or low vision. She has conducted several studies and experiences to understand visually impaired abilities and behaviors and design accessible systems that help them interact easily with technology. She has also published several articles on accessibility and authentication mechanisms for visually impaired users. She has also published articles related to using NFC technology and machine learning to enhance the healthcare systems.

MONA MOHAMMED ALNAHARI received the master's and Ph.D. degrees in information technology in USA, scholarly pursuits have centered around the realms of usability, accessibility, security, human–computer interaction, and AI. She is currently an esteemed Assistant Professor with Taif University, Saudi Arabia, stands as a beacon of knowledge and expertise in the field of information technology.

NAVEED EJAZ is currently a Postdoctoral Fellow with the GOAL Laboratory, School of Computing, Queen's University, Canada. His main research interests include image processing, computer vision, natural language processing, and machine learning. His expertise extends to applying these technologies in video and image analysis domains. Before joining the GOAL Laboratory, he worked in various academic and industrial roles in multiple countries worldwide.

REEMIAH MUNEER ALOTAIBI received the B.Sc. degree from Imam University and the M.Sc. and Ph.D. degrees from Leeds Beckett University, U.K. She is currently passionate about information management, big data, data governance, and user behavior. She is also an Assistant Professor with the College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Saudi Arabia, where she practices skills research and teaching. She has authored articles published in scientific journals, book chapters, and conference materials; and has also acted as a Reviewer of some journals, such as *International Journal of Civic Engagement and Social Change* (IJCESC) (IGI Global).

• • •