International Journal on Document Analysis and Recognition (IJDAR) (2022) 25:265–279
https://doi.org/10.1007/s10032-022-00412-9

SPECIAL ISSUE PAPER

BCBId: first Bangla comic dataset and its applications

Arpita Dutta[1] · **Samit Biswas**[1] · **Amit Kumar Das**[1]

Received: 15 March 2022 / Accepted: 22 August 2022 / Published online: 15 September 2022

Abstract

Comic document image analysis is now an active field of research in both academia and industry. However, comic document image processing research suffers due to its inherent complexities and the limited availability of benchmark public datasets. This paper describes the creation of the first-ever comic dataset among Indian Languages, namely *Bangla* Comic Book Image dataset *(BCBId)* (https://sites.google.com/view/banglacomicbookdataset), which is also made public for the benefit of the researchers. *BCBId* consists of 3327 images taken from 64 *Bangla* comic stories written by 8 writers. *Bangla* is the 6th most popular spoken language in the world—used by 265 million people (https://en.wikipedia.org/wiki/Languages_of_India), and has a century-old heritage of comic strips (in newspapers) and books. BCBId has the ground truth for extracting various visual components of the comic book images, i.e., panels, characters, speech balloons, and text lines. BCBId also includes the metadata encoding of all images in XML format to describe the underlined structure, semantics, and other features of the documents to pursue research on understanding stories and dialogues. A tool is specifically designed for accurate and faster ground-truth generation. As an application of the dataset, we carry out the sentiment analysis of comic stories—the first-ever attempt on comic book images. We also elaborate on a couple of applications of the BCBId in the comic research domain. Besides, we estimate the errors made by the annotators during the annotation process and describe different evaluation parameters to test the efficacy of the comic document image analysis algorithms.

1 Introduction

Over the past century, a number of countries have produced many comic books since comics are mesmerizing visual texts that create long-lasting impressions among a huge number of audiences, from eight to eighty. Comic artists incorporate various components such as panels, speech balloons, text-boxes, characters, gutters, and narrative text boxes to illustrate the entire story aesthetically (Fig. 1). Research works are being carried out on comic books to provide a fantastic user-friendly experience to the readers while reading them digitally. These researches focus on multiple aspects, such as interactive reading [3,29,52], content adaptation [27], retrieval of comics [33], development of new technologies for visually challenged people [11], and automatic content understanding [35]. In a nutshell, the primary targets of all these research works are—(1) extraction of various elements from comic images, (2) analysis of the relationship between these elements, and (3) the emotion/sentiment analysis of comic scenes. Some research works have focused on the first two challenges, i.e., automatic extraction of various components of comic images and analysis of the relationship between them [5]. However, sentiment analysis of comic scenes and automatic content understanding are still open areas of comic research that need strong research efforts from the community. The objective of the present work is to help researchers in this field by mulling more comic images available with adequate and authentic ground truth. We also believe that the detailed description of the development of this database and ground truth generation would give a lot of insights to the comic document image processing community. To our knowledge, there is no publicly available dataset of comic books in any Indian language. So, we decide to add this dataset on Bangla comics, considering the rich heritage and literary trait of Bangla among the Indian Languages. Moreover, the total volume and varieties of Bangla comics

✉ Arpita Dutta
arpita_dutta.rs2018@cs.iiests.ac.in

Samit Biswas
samit@cs.iiests.ac.in

Amit Kumar Das
amit@cs.iiests.ac.in

[1] Department of Computer Science And Technology, Indian Institute of Engineering Science and Technology, Shibpur, India

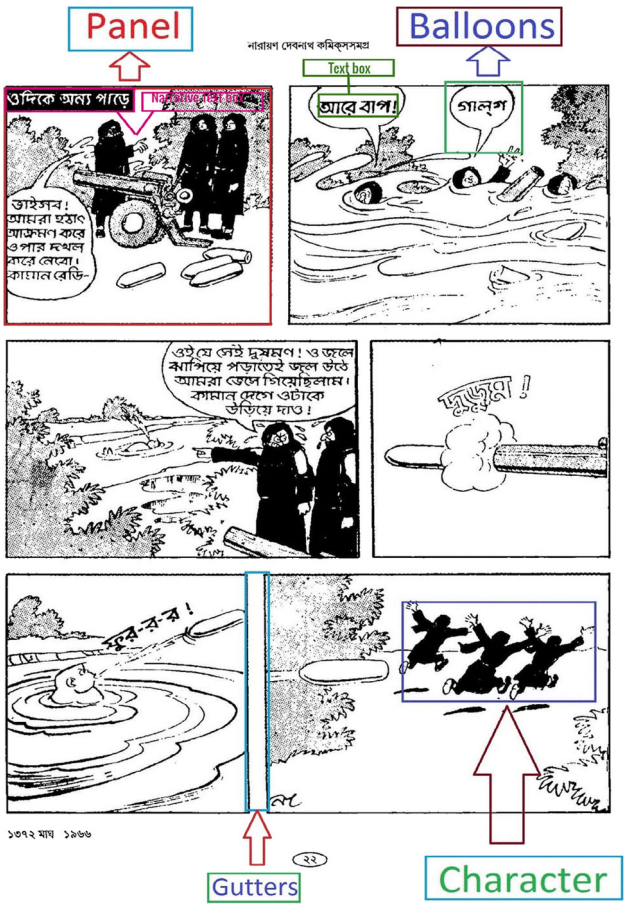

Fig. 1 Visual components of comic (Source:BCBId)

from many famous writers and easy availability are also the factors for the choice of the language as well.

To sum up, in this paper, the two significant contributions of our research works are presented: (1) the creation of a new comic dataset along with the development of a new GUI-based annotation tool and the presentation of an efficient strategy to approximate the errors caused by the annotators during the annotation process; (2) sentiment analysis of comics based on textual content and demonstrations of experimental results of various analysis task of comic evaluated on this dataset and other publicly available datasets along with the description of some evaluation parameters.

2 Related work

Here, our discussion of the previous research works is primarily based on (1) visual element extraction and (2) sentiment analysis-related aspects.

2.1 Visual element extraction

In [5], the readers may get beautiful insights about various comic related research works. Most of these works focused on the automatic extraction of various component of comic book page images such as panel extraction [12,20,37,45,51], speech balloon extraction [11,13,34,44,46] and comic character extraction [12,32,34,39,48]. However, a limited number of comic document image datasets are publicly available for pursuing academic research in this field due to copyright issues. The four publicly available comic datasets are eBDtheque, Manga 109, DCM772, and COMICS. A brief summary of these four publicly accessible comic datasets is presented in Table 1.

Guerin et al. [17] created eBDtheque, the first publicly released comic dataset. This dataset contains a total of 100 pages of American, French, and Japanese comics. Each of these pages, taken from various sources, includes the details of its source information. Along with the position of panels, speech balloons, and text-lines, this dataset also contains different speech balloons of various shapes (cloud, typical, spike), the reading order of panels, direction of the tail of speech balloons, and so on. COMICS, Iyyer et al. [23], includes 3498 volumes of American comics taken from the Digital Comics Museum. This dataset contains the annotations of bounding boxes around panels and text. However, only 500 pages are manually annotated. A Faster R_CNN [42] is applied to annotate the rest of the images. Another available comic dataset is Manga 109 [1,28,36] which contains 21,142 images of Japanese comics written by 94 various authors. This dataset includes the annotations of characters, items, scene texts, speech balloons, panels, thoughts, and narration. On the other hand, the DCM772 dataset [33] consists of 772 images of 27 golden-age comic books, which are collected from the Digital Comic Museum [9].

2.2 Sentiment analysis

Most of the previous research works of sentiment analysis depend on the analysis of the textual contents. All the existing methods of lexicon-based approaches are categorized into three groups [6]—(1) statistical based, (2) knowledge based, and (3) hybrid approach. The NLP-based sentiment analysis techniques are further partitioned into two categories : (1) traditional based and (2) machine learning based. Most of the traditional-based approaches are based on lexicon sets [22,30,31] and parse tree or n-gram features [8,24]. Recently, the performance of sentiment analysis has been improved a lot with the evolution of different neural network algorithms [4,10,26,38]. Yadav et al. [53] beautifully reviewed the existing deep learning models of sentiment analysis.

However, these neural network models require a large amount of textual data for training purposes, which is a con-

Table 1 Some publicly available comic datasets

Dataset	DPI range	Document language	Year	Books	Pages	Annotation type
eBDtheque [17]	72–600	English, French Japanese	1905–2012	25	100	Panels, balloons characters, text-lines
Manga 109 [1]	100–300	Japanese	1970–2010	109	21,142	Panels, text-boxes characters (face + body)
DCM772 [33]	Varying	English	1938–1954	27	772	Panels, characters (face + body)
COMICS [23]	Varying	English	1938–1954	3948	198,657	Panels, text-boxes

Table 2 The details of the proposed *Bangla* Comic Book Image dataset *(BCBId)*

Writer name	Story name	Number of pages	Dimension	Total no. of pages
Narayan Debnath	Narayan Debnath Somogro (All stories)	509	845 × 1098	509
Shibram Chakrabarty	01Dakatar daklen Harshabardhan	24	2226 × 3072	
	02 Harshabardhan banam kalka kashi	12	2550 × 3300	
	03 Change e gelen Harshabardhan	12	2550 × 3300	
	04 Chor banam Harshabardhan	12	2550 × 3300	
	05 Harshabardhaner bon mohotsyab	12	2550 × 3300	112
	06 Chor dhorlo Gobardhan	12	600 × 828	
	07 Mao dhora ki sohoj naki	12	2550 × 3300	
	08 Ekalabyaer Mundupat	12	7500 × 10,000	
	09 Harshabardhan er bagh sikar	12	2550 × 3300	
Frank Cho (*Bangla*)	01 Bonkonna part1	25	2560 × 3840	
	02 Bonkonna part2	22	2560 × 3840	
	03 Bonkonna part3	24	1275 × 1651	95
	04 Bonkonna part4	24	640 × 953	
Satyajit Ray (Series 1)	01 Darjeeling Jomjomat	59	3450 × 4871	
	02 Gangtok Gondogol	79	2244 × 3045	
	03 Gosaipur Sorgorom	62	2253 × 3087	
	04 Jahangirer Swarnomudra	42	3204 × 4271	
	05 Nepolianer Chithi	43	3204 × 4271	496
	06 Robertsoner Rubi	48	675 × 911	
	07 Royal Bengal Rohossya	78	3200 × 4267	
	08 Seal Debota Rohossyo	37	3200 × 4267	
	09 EbaarKandoKedarnath	48	2480 × 3508	
Satyajit Ray (Series 2)	01 Aadim Manush	25	3750 × 5000	
	02 Manro dip r rohosso	32	3200 × 4267	
	03 Aschorjontu	30	3200 × 4267	
	04 Moru Rohosya	27	4129 × 5846	
	05 Prof. Sonku o robu	31	3200 × 4267	218
	06 Prof. Sonku o Aschorjo putul	31	3200 × 4267	
	07 Prof. Sonku o Chi ching	21	3750 × 5146	
	08 Prof. Sonku o Har	21	3792 × 5117	
Premendra Mitra	01 Chhori	73	12,500 × 15,417	
	02 Mosha	15	4959 × 7017	
	03 Nuri	17	12,500 × 15,000	137
	04 Poka	32	12,500 × 17,500	

Table 2 continued

Writer name	Story name	Number of pages	Dimension	Total no. of pages
Hergé_Tintin (*Bangla)*	01 Congo'y TinTin	64	2567 × 3613	
	02 America'y TinTini	64	3333 × 4142	
	03 Pharao'er Churut	64	3333 × 4400	
	04 Nil Kamal	64	2550 × 3300	
	05 Kanbhanga Murti	64	3333 × 4692	
	06 Otokar'er Rajdando	64	3333 × 4583	
	07 Kankra Rahoshyo	64	6667 × 9167	
	08 Ashchorjo Ulka	64	6667 × 9042	
	09 Bombete Jahaj	64	6667 × 8892	
	10 Lal Bombete'r Guptodhan	64	6667 × 9033	
	11 Mommy'r Abhisap	64	6667 × 9133	1216
	12 Suryadeber Bandi	64	5063 × 6845	
	13 Kalo Sonar Deshe	64	5087 × 6518	
	14 Chandraloke Abhijan	64	6667 × 9167	
	15 Chande TinTin	64	6667 × 9167	
	16_Tibbot_e_Tintin	64	4958 × 7017	
	17_Biplabider_Dangole	64	6667 × 8842	
	18_Flight_714	64	6667 × 9342	
	19 Calculus er kando	64	5100 × 6600	
Indrajal	01 Ajob Desher Bondi	32	1400 × 2002	
	02 Aleyar Hathchhani	32	1400 × 1948	
	03 Ghomta Dhaka Rohossyo	36	1696 × 2422	164
	04 Him Nissas	32	2550 × 3489	
	05 Othoi Joler Guptodhon	32	1400 × 1956	
Mandrake (*Bangla*)	01 Countdown to Oblivion	80	5383 × 1625	
	02 Tollbooth er Atonko	72	5383 × 1617	
	03 13 No. Cell er Bondi	86	5358 × 1600	444
	04 Atonko	93	5408 × 1608	
	05 Mountain of Mystery	113	5358 × 1592	

straint for comics due to copyright restrictions. Therefore, we propose a lexicon-based approach for the sentiment analysis of comic stories based on textual data.

3 BCBId

The preliminary version of this BCBId dataset is briefly presented in [13]. In this version of the BCBId dataset, we have added 1954 more images along with 25 more stories. A new type of XML annotation (CBML-based XML annotation, Sect. 4.2), which encodes the underline structure and semantics of comics stories, is also added in this version of the BCBId dataset. Furthermore, the detailed annotation procedure, the developed GUI-based annotation tool, and the error rectification strategy during the annotation procedure are also included here. Now, the entire BCBId dataset is the collection of 3327 images taken from 64 *Bangla* comic stories designed and animated by 8 writers. This entire dataset includes images, a wide variety of dimensions, and drawing styles used by comic artists. The BCBId also contains some *Bangla* translation of *English* comics like *Tintin*, *Mandrake*, and *Jungle Girl*. The details of the whole dataset are given in Table 2. The following subsections show the details of the *BCBId* Dataset and its ground truth.

3.1 Dataset details

This *Bangla* Comic Book Image dataset *(BCBId)* has the following attributes: (1) The images' resolution remains constant for a single story, and it changes if the story changes. (2) Most of these comic books are written in the twentieth century. Some of them include *Bangla* translation of *English* comic books, i.e., *Tintin*. (3) Most of the images are col-

lected from different *Bangla* webcomic websites, [1] which host user uploaded scan copies of several *Bangla* comics. (4) 80% of those images are color images. The rest are gray and binary images. (5) Each comic book page image contains fundamental elements of comic book like panels, characters, speech balloons, narrative text boxes, gutters, etc. (6) The shapes of the panels are heterogeneous. Most of the panels are regularly shaped closed panels. A significant amount of panels are open. Their boundaries are interpreted by the color difference between panels and the background. Moreover, some overlapping panels also exist, i.e., those panels are not separable by using boundaries. (7) The shapes of the speech balloons are mostly oval, rectangular, peaky, wavy, circular, and square with white backgrounds. Most of them have a tail pointer, which links them to their respective speakers. (8) Texts are mostly handwritten; some of them are printed. The text portion mainly appears within speech balloons and narrative text boxes. However, some text appears neither in speech balloons nor in narrative text boxes, and those texts are mainly used to express the corresponding contextual emotions.

4 BCBId annotations

Although ground truth is an essential part of evaluating the results of any algorithms and pursuing further research, developing those ground truths is always a tedious task. The GUI (graphical user interface)-based annotation tools reduce the annotator's efforts and speed up the overall annotation process. For pursuing the two main aspects of comic research, the entire annotation procedure of the BCBId dataset is conducted in two different steps: (1) content annotations and (2) CBML-based XML annotations. Table 3 demonstrates the detail descriptions about both content annotations and CBML-based XML annotations.

4.1 Content annotations of comic images

This research domain focuses on extracting and analyzing various contents of comic images such as panels, characters, text boxes, speech balloons, and narrative text boxes. Though this task is similar to visual scene understanding, the huge variety in drawing style and lack of publicly available ground truth data makes the task more challenging than natural images. Each of those components is annotated using the following guidelines:

(1) *Panels* In comic book images, an image area demonstrating a single scene is called a panel. There must exist at least one panel on a page. The bounding boxes are drawn as close as possible across the boundaries to trace the closed

[1] http://www.bengalicomics.com/p/home.html.

Table 3 Details of Content-based and CBML-based XML annotations of the BCBId Dataset

Writer	Annotated stories	Content annotated pages	Panels	Characters	Speech balloons	Text boxes	Narrative text boxes	Balloon shape	XML annotated pages
Narayan Debnath	3	120	350	420	510	707	197	Irregular	310
Shibram Chakrabarty	7	107	210	359	337	547	210	Rectangular	276
Frank Cho	4	98	312	401	521	996	475	Elliptical	135
Satyajit Ray	14	522	824	920	1135	1664	529	Rectangular	970
Premendra Mitra	4	210	437	569	749	1077	328	Elliptical	341

panels. In the case of open/frame-less panels, the bounding boxes are drawn depending on the understanding of the scene. (2) *Balloons* The speech balloons represent the conversation between characters, which helps the readers understand the comic content and follow the story pattern. In the case of closed balloons, the boundaries are located first, and then they are annotated at the pixel label. To deal with open speech balloons, the boundaries are approximated first and then annotated at pixel label. (3) *Text boxes* The bounding boxes, named text boxes, are drawn to locate the text region inside the speech balloon and narrative text boxes. Once the text boxes are appropriately located, the inside text can be used for other research purposes like text-line segmentation, character recognition, OCR, etc. Moreover, this text is encoded with more intricate details for other research purposes. (4) *Comic Characters* One of the essential tasks of comic content analysis is to extract the characters of comic books. The characters are located by drawing bounding boxes around them (including face and body together). (5) *Narrative Text Boxes* The narrative text boxes, also referred as captions, are used to represent narrative text. The bounding boxes are drawn to locate these regions.

(A) *Content Annotation Procedure* The drawing of these visual components is entirely independent of each other. For example, it is not necessary that a speech balloon or a character must appear within a single panel. Therefore, each of these components should be annotated separately in order to detect them easily from the images. Manga 109 [1] developed an annotation tool to provide bounding box annotations for the panel, character faces, character body, and text. However, the content annotations of the BCBId dataset contain polygonal masks annotations for speech balloons and bounding box annotations for the panel, text boxes, characters, and narrative text boxes. Therefore, with the help of a well-established publicly available VGG ANNOTATION TOOL [14,49], we annotate each of those components separately and export those annotations into plain text formats (CSV and JSON) to utilize them for further processing.

(B) *Error calculation* The outcome is error-prone when many people are engaged in developing such types of content-based annotations. We introduce an error calculation strategy only to rectify the errors being incurred while developing content-based annotation. The position of corner points of the bounding boxes around panels, characters, and text lines is crucial while preparing the ground truth for visual segmentation. Instead of utilizing a particular number of pixels, a certain percentage of the entire page size is used to fix the error of the corner points. In our case, the percentage P is set to the 0.35% of the page width and height on the y and x axes, respectively. For example, if the test image size is 1690×2195, this creates a delta of +/– 7 pixels along the x-axis and +/– 6 pixels along the y-axis, respectively.

Fifteen people have been engaged in developing our dataset. Each of these fifteen persons is asked to draw the four corner points of bounding boxes. Then, the mean position from fifteen values is computed for each corner point around bounding boxes. Then, the distance of each of those points to its mean is calculated. After analyzing the data, it has been observed that the 90% pointed corners have a distance of 5 pixels to mean position for the threshold value of P set at 0.35. In the case of huge improper segmentation, the errors are corrected manually. This error criterion strategy is successfully applied to panels, characters, and text lines. Although speech balloons have polygonal boundaries, this strategy gives promising outcomes.

4.2 CBML-based XML annotations of comics

Another research area on comic book images focuses on understanding the story and dialogue of comics depending on natural language processing. Therefore, the information on each page should be encoded in such a way so that it can be processed later to identify the underlined structure, semantic, and other features of the document. The semantic annotation of a given page is stored as XML encoding, which includes parts of CBML (Comic Book Markup Language) [50] with some more additional tags. This XML-based format provides essential benefits to the database since the predefined metadata element helps to encode the information of a page in an easier and faster way.

The complete descriptions of comic book pages are encoded with the help of a total of 13 tags. The XML encoding begins with the root element <panelGrp>, which binds the information of all panels of a single page together. Each root element contains several <panel> children, which are used to describe the complete information related to each panel. Each of these <panel> node is described through some attributes $characters$, n and id. Here, the attribute, $character$, signifies the total number of characters present in that particular panel along with their name. The attribute, n, signifies the rank of the panel on that page. The attribute, id, is used to identify the position of that panel within the page. For example, the first panel of a comic page is denoted by <panel characters=" #1 Cap #2 Batul " n="1" id="eg_000">. Here, $character = "\ \#1Cap\#2Batul"$ signifies that the first character is Captain and the second character is Batul in that panel. $n = 1$ denotes the first panel of that page, and $id = eg\_000$ signifies the panel exists in the first row of that page. Under <panel> node, the information related to speech balloons is described through <balloons> children node along with various attributes. Each of these <balloon> node is described through some attributes $type$, who and id. Here, the attribute, $type$ signifies the category of speech balloon, i.e., "speech," "telepathic," etc. The attribute, who, denotes the speaker of the speech balloon. On the other

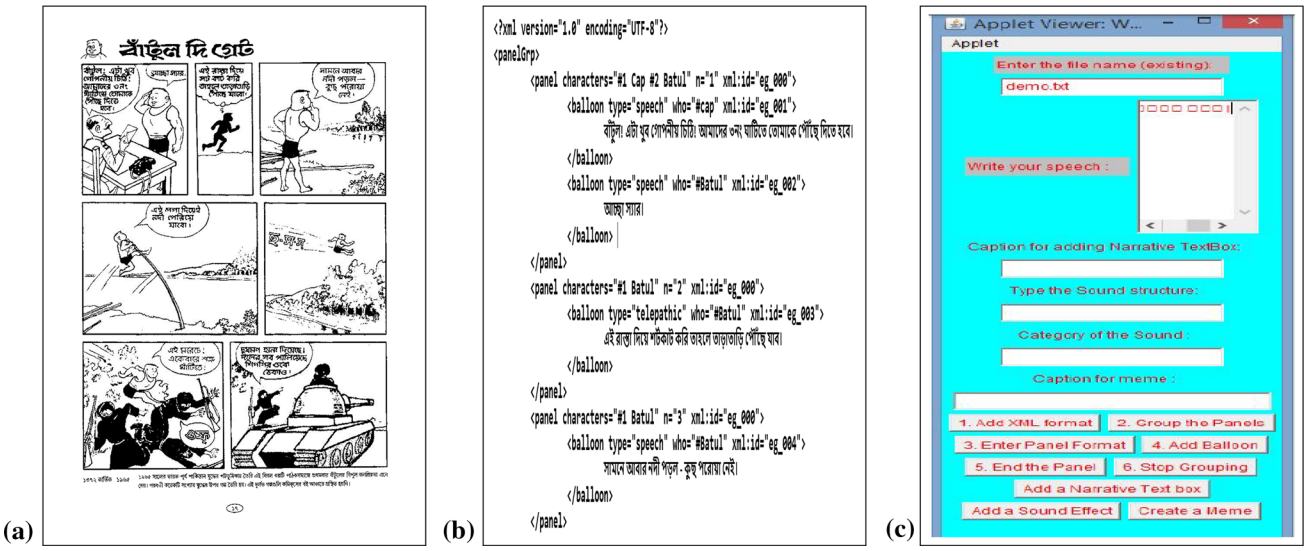

Fig. 2 **a** Input image; **b** Screenshot of the XML-based annotation up to third panel; **c** Developed tool

hand, the attribute, id, signifies the rank of speech balloons within comic pages. For example, the first speech balloon of a comic page is represented as <balloon type="telepathic" who="#cap" id= "eg_001">. Here, $type = '' telepathic''$ signifies that the type of the speech balloon is telepathic. The attribute, $who = '' \#cap''$, signifies that the speaker of this speech balloon is Captain. The attribute, $id = eg\_001$ signifies the first speech balloon of that page. The text within the speech balloon is then written under <balloon> node. Sometimes, only a single character is used within speech balloons to materialize some emotions and expression (i.e., '?' is used to express a lack of understanding or '!' for expressing surprise). A single word is also used to express certain emotions. Another child <emph> under <balloon> node is used to encode these special types of emotions and expressions to distinguish those from normal text speech. However, some texts, which do not appear within any speech balloon, are used to describe certain emotions (i.e., especially graphic sound). That information is encoded separately under <sound> node within each panel if such sound exists. The narrative texts, i.e., captions of panels, are written under <caption> node. The <caption> node has an attribute id which identifies the position of the caption, i.e., $id = eg\_001$ signifies the first caption of that page. The metadata of that page, i.e., page number, title, etc., are also recorded under <fw> node. The value of the attribute $type = ' pageNum'$ records the page-number of that page. The input image and its corresponding CBML-based XML annotations are depicted in Fig. 2. A brief summary of the tags introduced is given in Table 4.

Developed tool The creation of the CBML-based XML annotation is always a tedious and time-consuming task, and it needs a considerable amount of effort. There exists no available GUI-based annotation tool for creating such kinds of CBML-based XML annotations. We developed a GUI-based ground truth construction tool, coded utilizing Java and Java's Applet web browser to take advantage of platform independence and reduce the server's burden. The user can efficiently encode the information of a comic book page in an interactive way by using the tool, and it automatically generates the corresponding information in XML annotation format. Moreover, the designed tool also provides an elegant display of the unlined XML tree structure and semantic relationship between the various elements of the comic book page. This tool also facilitates the user to encode the textual information in any language according to their need. Nine various buttons control this annotation tool—(1) Add XML format, (2) Group the panels, (3) Enter Panel format, (4) Add Balloon, (5) End the panel, (6) Stop grouping, (7) Add a Narrative Text box, (8) Add a Sound effect, and (9) Create a Meme, as depicted in Fig. 2c. Each time a button is pressed, the corresponding XML structure along with the designated attributes (as described in Sect. 4.2) will be embedded automatically into the annotation file. For example, the XML format <balloon type="speech" who="#Batul" id= "eg_001"> is added automatically into the annotation file by pressing "Add Balloon" button and selecting the corresponding attribute values. The text fields, namely, "Write your speech" and "Caption for adding Narrative TextBox," include speech texts and narrative texts within balloons and narrative text boxes, respectively. Moreover, the text field, namely "Category of the sound," assigns the emotion labels (such as excitement, anger, sadness, happiness, surprise, etc.) associated with the corresponding sound. Such types of emotion labels are incorporated in a separate CSV file for utilizing them further in several other applications like comic style video summariza-

Table 4 A brief description of introduced tags

No.	Tags/variables	Format	Description
1	PanelGrp	<panelGrp>...< /panelGrp>	**panelGrp** tag is used to bind together all the panels fitted and confined to that particular page only. It denotes that there are more than one panel in that particular page
2	Panel	< panel>...< /panel>	**panel** tag indicates the starting of a panel in a particular fragment
3	Characters	characters="#1 name1 #2 name2 "	**characters** indicate the number of characters that are present in that particular panel. For indicating the count of the characters '#1'tag is used. This will increase as the number of characters will increase
4	n	n="1"	**'n'** denotes the number of the panel in that particular page, i.e., n="1"means first panel of that page
5	xml:id	xml:id="eg_000"	It is used to denote position of the element
6	Balloon	< balloon>...< /balloon>	**balloon** tag indicates the starting of a phrase or sentence that is to be said by a particular character
7	Type	type="speech"	**type** specifies the category of balloons, which can be a 'speech"or a 'telepathic"one
8	Who	who="#name"	It specifies the name of the character who is speaking
9	Place	place="middle"	**place** denotes that the page number is in the middle position of the page
10	Type	type="pageNum"	**type** is used for denoting the page number
11	Fw	< fw>...< /fw>	It is used to denote the end of a particular page
12	Caption	< caption>...< /caption>	**caption** is used to denote the caption of a particular panel
13	Sound	< sound>...< /sound>	**sound** tag is used to specify the sound effect

tion [16], comic video generation [18], etc. Furthermore, this annotation tool also gives the privilege of annotating social media memes by utilizing the "Create a Meme" button. This tool is publicly available[2] for research purposes with a user manual as a complete annotator guideline.

5 Various applications of *BCBId* dataset

The proposed dataset has been developed with the focus on pursuing research work into two different domains. One of them is associated with the extraction of various components, i.e., panels, characters, speech balloons, and narrative text boxes from a comic book page. In contrast, the other research domain focuses on analyzing the encoded metadata of comic pages for the intuitive understanding of comic stories, sentiment analysis of comic stories, etc., based on Natural Language Processing.

5.1 Visual components extraction from comics

The primary objective of this research work is to detect various components of comic images such as panels, characters, text boxes, speech balloons, and narrative text boxes. However, there exists almost no structural homogeneity among the layout structures of these multiple components of comic images.

5.2 Sentiment detection of comic stories

The primary focus of this research area is the prediction of sentiment associated with comics which plays a significant role in some other aspects of comic research like intuitive understanding of comic stories, genre prediction of comics, etc. To pursue research in this domain, we encode all the semantic information of a comic book page into XML format with various tags (see Table 4). In this paper, the proposed method detects the contextual sentiment associated with the textual part of comic stories depending on Natural Language Processing. In this work, we consider a total of three sentiment labels. They are positive, negative, and neutral. We analyze the raw text information under **balloons** and **caption** tags (see Table 4) as the speech balloons

[2] https://drive.google.com/drive/folders/18g2GeLvmoXJLMhk27zzs30rQxxUpgUS5?usp=sharing.

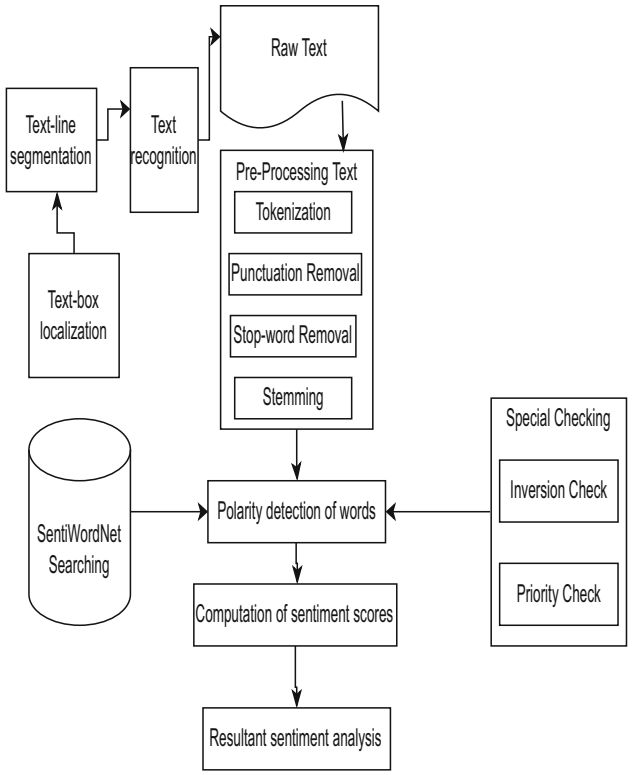

Fig. 3 Brief overview of the proposed method

and captions are the text-containers of comic stories. The raw data extraction precede the following three steps: (1) text-box localization, (2) text-line segmentation, and (3) text recognition. We perform text localization, text line segmentation, and text recognition by utilizing Dutta et al. [12], Dutta et al. [13], and Hartel et al. [19], respectively. The proposed method of sentiment detection is language independent. Since *English* is an international language, we explain the proposed approach explicitly with the examples given in both *Bangla* and *English*. Figure 3 depicts a brief overview of our proposed algorithm.

(A) *Pre-Processing Text* The raw text information extracted from different comic book pages contains some unwanted and noisy data. Therefore, this information should be processed properly in order to obtain clean and suitable text data. Then, those clean data are analyzed in order to retrieve the proper sentiment associated with it. The following steps are taken for the pre-processing of the raw text data.

(1) *Tokenization* Tokenization is the process where a text line is partitioned into several components like phrases, words and symbols. Each of these components is called a token. For example, the following text line is taken from one of the comic pages of the BCBId dataset:

কাজটা আমি খুব চমৎকার করতে পারি ! আপনি অনেক আনন্দের অপেক্ষায় থাকুন।

The ***English*** translation of the above ***Bangla*** text is—***I can do this job very well ! You just wait for much enjoyment.*** After tokenization, the output for the *Bangla* text will be—

'কাজটা', 'আমি', 'খুব', 'চমৎকার', 'করতে', 'পারি', '!', 'আপনি', 'অনেক', 'আনন্দের' , 'অপেক্ষায়', 'থাকুন', '।'

Similarly, after tokenization, the *English* text will produce the following output—'I', 'can', 'do', 'this', 'job', 'very', 'well', '!', 'You', 'just', 'wait', 'for', 'much', 'enjoyment', '.'

(2) *Punctuation Removal* Raw text data may contain some punctuations that have very little impact on sentiment analysis. Therefore, those punctuation are removed at pre-processing step. After punctuation removal, the output for the *Bangla* text will be—

'কাজটা', 'আমি', 'খুব', 'চমৎকার', 'করতে', 'পারি', 'আপনি', 'অনেক', 'আনন্দের', 'অপেক্ষায়', 'থাকুন',

After punctuation removal, the output for the *English* text will be—'I', 'can', 'do', 'this', 'job', 'very', 'well', 'You', 'just', 'wait', 'for', 'much', 'enjoyment'

(3) *Stop-word Removal*Stop words are used to represent a sentence in a better way, but they have no impact on sentiment analysis [25]. Therefore, we scan all the tokens at the pre-processing step and remove the tokens that match the list of stop words. Some examples of stop words in *English* language are—'I', 'You', 'We', 'Our', 'just', 'can', 'do', 'so', 'therefore', 'and', 'this', 'that', etc. Some of the stop words in *Bangla* are

'আমি', 'আপনি', 'আমার', 'করি', 'থাকুন', 'পারেন', 'পারি', 'করতে', 'করেন', 'এবং', 'অতএব', 'যাহাকে', 'সুতরাং',

etc.

After stop-words removal, the output for *Bangla* text willbe 'কাজটা', 'খুব', 'চমৎকার', 'অনেক', 'আনন্দের', 'অপেক্ষায়'

The output for *English* text will be—'job', 'very', 'well', 'wait', 'much', 'enjoyment'

(4) *Stemming* In the raw text, a single word may appear in many forms, but the original/root word is needed to identify the sentiment associated with it. The lexicon dictionary stores the original/root form of the sentiment words with their corresponding sentiment score. Therefore, first, all the tokens are scanned for stemming. In *English*, the words ending with 'ious', 'ment', 'ness', 'ing', 'ly', etc., are stemmed to extract the corresponding root words. In *Bangla* text, if a word ends with া, ি, ে, ী, 'র', 'ই', 'য়', 'টি', 'ট', etc., then that word is stemmed. After stemming operation, if any match occurs with the lexicon dictionary, then that token is replaced with the stemmed word from the dictionary. Also, the associated sentiment score is retrieved to be analyzed later. In contrast, the tokens, for which, the associated stemmed words

do not match with the dictionary, are kept as it is. After stemming, the final token list for the *Bangla* text will be—'কাজটা', 'খুব', 'চমৎকার', 'অনেক', 'আনন্দ', 'অপেক্ষায়' Note that, here, ('আনন্দের') is the sentiment word that is replaced with stemmed word ('আনন্দ') However, the tokens ('কাজটা', 'অপেক্ষায়') remain constant as they are not sentiment words. Similarly, after stemming, the final token list for the *English* text will be—'job', 'very', 'well', 'wait', 'much', 'enjoy.' Like *Bangla* text, here also, the sentiment word ('enjoyment') is replaced with the stemmed word ('enjoy').

(B) *Polarity detection of words* In the proposed method, three categories of sentiments, i.e., positive, negative and neutral, are considered. After pre-processing of raw text information, those cleaned text data are matched with the lexicon dictionary SentiWordnet [7,15]. Then, each of these matched words is assigned a polarity (positive, negative, and neutral) according to the SentiWordnet [7,15].

Special Checking There exist some words which can change their polarities according to the context. Therefore, the associated sentiment scores of these words cannot be predicted directly. To deal with this, we check the another two following conditions before assigning sentiment score to a word.

(1) *Priority Check* We create a list of priority words that elevate the polarity of its next words. Therefore, the polarity of a word is decreased or increased according to the presence of these priority words in a text line. In *English*, some of the priority words from the proposed list are—'many', 'much', 'few', 'very', 'most', etc. In *Bangla*, The proposed list contains the priority words such as 'খুব', 'সবচেয়ে', 'অনেক', 'অল্প', 'অত্যন্ত', 'অত্যধিক',

etc. According to the list of priority words, in the *Bangla* text, 'খুব' and 'অনেক' are two priority words that elevate the polarities of the next two following words 'চমৎকার' and 'আনন্দ', respectively. Here, 'চমৎকার' and 'আনন্দ' are two sentiment words with positive polarities. Therefore, the words 'খুব' and 'অনেক' increase the polarities of the text line. Similarly, in the *English* text, 'very' and 'much' are two priority words that elevate the polarities of the next two following words 'well' and 'enjoy', respectively. Here, 'well' and 'enjoy' are two sentiment words with positive polarities. Therefore, the words 'very' and 'much' increase the polarities of the text line.

(2) *Inversion Check* Sometimes, the presence of inverted words changes the contextual meaning of a sentence. In *Bangla*, some of the inverted words are 'নয়', 'না', 'নি','নাই', etc. Similarly, in *English*, some examples of inverted words are 'no', 'not', 'never', 'neither', etc. Therefore, in our method, if any inverted word is found within a sentence, -1 is multiplied with the sentiment scores associated with the sentiment words to invert the polarity of the sentiment (i.e., positive polarity will be converted to negative and vice versa). Let us consider the sample *Bangla* text—কাজটা আমি খুব খারাপ করি না।

The ***English*** translation of the above ***Bangla*** text is—***I do not work badly***. In the *English* text, 'bad' is a sentiment word with negative polarity, the presence of the inverted word 'not' makes it positive by changing the contextual meaning of the sentence. Similarly, in the *Bangla* text, tough, 'খারাপ' is a sentiment word with negative polarity, the presence of the inverted word 'না' makes it positive by changing the contextual meaning of the sentence.

(C) *Computation of sentiment scores* After the pre-processing of raw text and polarity checking, in this step, each final token is matched with the lexicon dictionary SentiWordnet [7]. Then, according to the lexicon dictionary SentiWordnet [7], the sentiment score for each final token is assigned a sentiment score based on the following rules :

(1) Each of the final tokens is assigned a sentiment score within the range -2 to $+2$. Here, the sentiment score for strong positive words is assigned as $+2$, and the sentiment score for strong negative words is assigned as -2.
(2) If a token does not match any sentiment word, that token is considered neutral and assigned a zero sentiment score.
(3) If the final token list contains any priority word, then the sentiment score of its next sentiment word is multiplied with 2 to elevate the score.
(4) If any inverted word presents, then the sentiment score of the sentiment word is multiplied with -1 to invert the context.
(5) Then, the total sentiment score for each sentence will be— $S_{\text{score}} = \sum_{i=1}^{n} S(T_i)$. Here, S_{score} represents the total sentiment score computed for a sentence where $S(T_i)$ denotes the sentiment score associated with ith token T_i.

(D) *Resultant sentiment analysis* The S_{score} should be normalized [3] within the range -1 to $+1$ in order to predict the proper sentiment associated with a sentence. The normalized sentiment score for a sentence is now represented as $NS_{\text{score}} = \frac{S_{\text{score}}}{\sqrt{{S_{\text{score}}}^2+\alpha}}$, where $\alpha = 15$. Therefore, the normalized sentiment score NS_{score} now exists within the range -1 to $+1$. In our paper, if NS_{score} of a sentence is greater than -1 and less than -0.2 ($-1 < NS_{\text{score}} < -0.2$), it is considered as negative. The sentences having NS_{score} greater than 0.2 and less than 1 ($0.2 < NS_{\text{score}} < 1$) are considered as positive. The sentences with NS_{score} greater than equal to -0.2 and less than equal to 0.2 ($-0.2 \leq NS_{\text{score}} \leq 1$) are regarded as neutral.

[3] https://www.nltk.org/api/nltk.sentiment.html.

6 Evaluation parameters

Along with the database, some parameters are also provided for the evaluation of various kinds of algorithms of different tasks. Suppose, $E_x = (x_1, x_2 \ldots x_s)$ and $G_y = (y_1, y_2 \ldots y_t)$ are the set of s extracted and t ground truth, respectively, for same kind of elements (panels, speech balloons, textlines and characters). All the objects of E_s are compared with each of the elements of G_y, and the corresponding matching elements of E_s are labeled as correctly extracted objects.

(1) *Parameters for visual component extraction* In the case of panels, characters and text-line extraction algorithm, the bounding boxes having more than 80% overlap with their corresponding ground truth are considered as true positive, i.e., T_P. However, the extracted bounding boxes having no intersections with any of its ground truth boxes are taken as false positive, i.e., F_P. Also, those ground truth bounding boxes, which have no matches, are regarded as false negative, i.e., F_N. Therefore the p_r, r_c and F_m are defined as: (1) $p_r = \frac{T_P}{T_P+F_P}$, (2) $r_c = \frac{T_P}{T_P+F_N}$ and (3) $F_m = 2 \times \frac{p_r \times r_c}{p_r + r_c}$. Similarly, IoU is defined as $IoU = \frac{E_s \cap G_y}{E_s \cup G_y}$. For speech balloon extraction, shared pixels are used to define metrics as the annotations are in pixel labels. But in some cases, there may exist some mismatches in segmented features though visual features match perfectly. Instead of ignoring those elements completely, a new technique is used to decrease those elements' evaluation scores. The evaluation score is determined based on the introduction of a new function F. Firstly, any extracted element x_i is regarded as validated if the value of the precision and recall of that element with its corresponding matching element are, respectively, more than two thresholds t_{p_r} and t_{r_c}. Therefore, this element is given a score (S_c) based on the following properties: (1) One-to-one matching: If x_i is matched with exactly one element y_j of G_y, then $S_c = 1$. (2) One-to-many matching: If x_i is matched with more than one element y_j, i.e., a subset $G_y^{\wedge}$ of G_y, then $S_c = 1 - |F(G_y^{\wedge})|$, where $|G_y^{\wedge}| > 1$. (3) Many-to-one matching: If more than one x_i , i.e., a subset $E_x^{\wedge}$ of E_x, is matched with exactly one element y_j of G_y, then $S_c = 1 - |F(E_x^{\wedge})|$, where $|E_x^{\wedge}| > 1$. The function F is any nonlinear function and in our work, $F(x)$ is regarded as $\ln(x)$. Moreover, the thresholds values of t_{p_r} and t_{r_c} are not fixed to any particular values. Instead, the values are determined empirically.

(2) *Parameter for sentiment analysis task* Here also, precision, recall, and F_{score} are used for measuring the performance of the proposed method on sentiment/polarity detection of comic stories. Here, the sentences, that are correctly identified, are recognized as true positive, i.e., T_P. In contrast, the sentences, that originally belong to negative class but are incorrectly categorized as positive class, are recognized as false positive, i.e., F_P. The sentences, which originally belong to positive class but are incorrectly identified as negative class, are labeled as false negative, i.e., F_N.

7 Result analysis

The performance of other methods from the literature is evaluated on the BCBId dataset. We also discuss the strength and weaknesses of the literature methods for both content and sentiment analysis.

(a) *Result for content analysis* Few research works have already been done on the content analysis of comic book pages. In this paper, we show the performance of these algorithms on the proposed BCBId dataset. Figure 4a depicts the implementation of various panel detection algorithms on the BCBId dataset. Since most of the existing panel detection algorithms [37,47,51] depend on the analysis of connected components and division lines, those methods do not perform well in the absence of white margin or the presence of complex layout structures of panels. To deal with these problems, A CNN (convolutional neural network)-based deep learning method is proposed to predict the locations of panels directly from comic images [12]. Later, another similar kind of deep learning architecture like [12] is reported in [18] for panel detection.

Figure 4b depicts the performance of various algorithms on character detection. There are huge variations in the expressions, position of organs, shapes, and sizes of comic characters. Moreover, comic characters are sketched using curved lines or straight lines and have little color information. These two constraints make the comic character detection problem more challenging than human face detection from real images. Some of the existing algorithms for comic character recognition [45,48] rely on color descriptors and the similarity between various features such as facial expressions, poses, shapes, etc. With the recent advancement in deep learning, the deep neural networks proposed by [12,32,34,39–41] tackle this challenge more efficiently with the significant improvement in the overall accuracy of the character detection problem. In narrative text boxes and speech balloon detection problems, the main challenge lies in capturing the entire balloon along with its tail which appears in many different styles and shapes. The previous heuristic-based algorithms [2,21,43,44], that mostly rely on the bounding box alignment around the inside text of speech balloons, connected component labeling, and contour analysis, lack clarity and details. Although two recently proposed deep neural networks [11,34] improve the accuracy by considering it as a pixel-based semantic segmentation problem, their performance degrades in bright areas due to the improper prediction of confidence score. To overcome this difficulty, a dual-stream neural network architecture is proposed in [13] by amalgamating both edge features

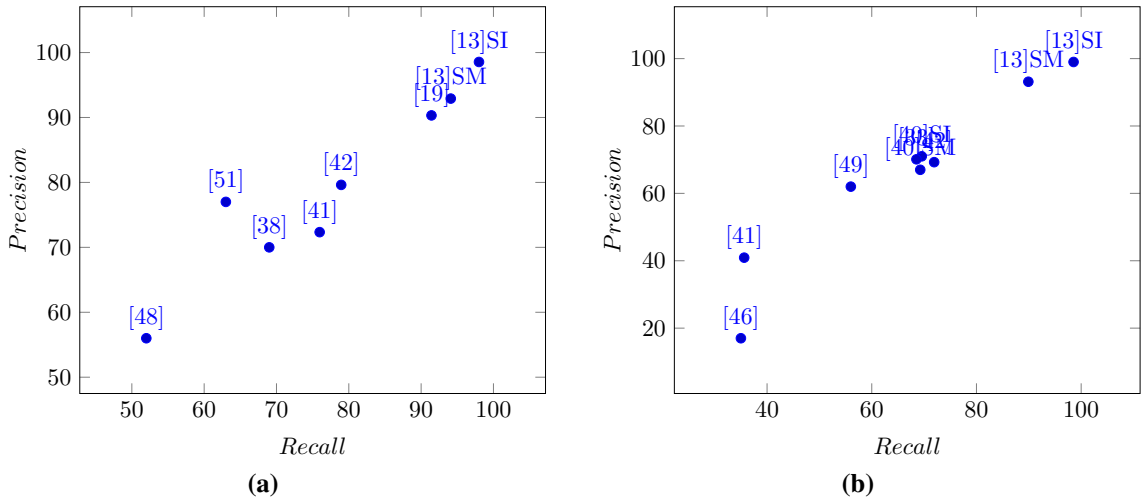

Fig. 4 Performance on BCBId dataset [13]: **a** different panel detection algorithms; **b** various algorithm on character detection. Here, SM stands for softmax function and SI stands for sigmoid

Table 5 Speech balloon detection on BCBId

Method	r_c	p_r	F_s
Arai et al. [2]	19.21	22.17	20.58
Ho et al. [21]	13.65	34.97	19.63
Rigaud et al. [44]	69.98	32.99	44.85
Rigaud et al. [43]	79.92	83.27	81.56
Dubray et al. [11]	89.89	91.64	90.81
Comic MTL [34]	91.32	92.27	91.80
Dutta et al [13]	97.05	98.81	97.92

Table 6 Narrative Text-box detection on BCBId

Method	r_c	p_r	F_s
Comic MTL [34]	90.38	91.22	90.87
Dutta et al. [13]	95.63	98.52	97.05

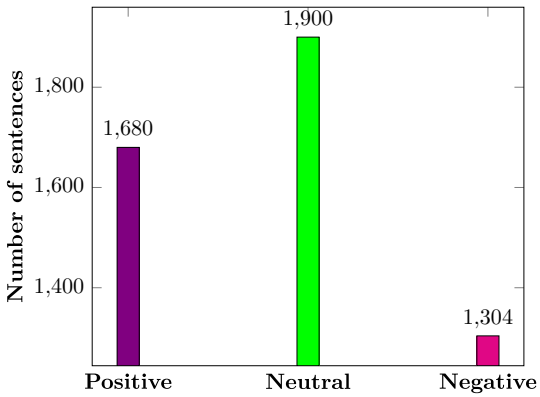

Fig. 5 Prediction of positive, neutral, and negative sentences by our method on BCBId dataset

and pixel-label semantics, and it successfully captures both speech balloons and narrative text boxes along with the huge variations in illustrator styles. However, the separation of overlapped speech balloons is still a challenging task. Table 5 demonstrates the performance of various speech balloon detection algorithms evaluated in the BCBId dataset. Similarly, Table 6 demonstrates the performance of narrative text box detection evaluated in the BCBId dataset.

(b) *Result for sentiment analysis* We encode the textual descriptions of all the contents of almost 2032 comic book pages from the BCBId dataset (Table 3) with the help of various tags (Table 4). In our dataset, there exist a total of 4884 sentences under the tags **balloons** and **caption** that are analyzed for the sentiment prediction of comic stories. Figure 5 depicts the sentiment predicted by the proposed system on the BCBId dataset. The result of Table 7 demonstrates some instances of texts from the BCBId dataset and their predicted sentiment by our proposed approach. Although several lexicon-based tools are available to evaluate the performance of lexical texts in *English*, very few are available for *Bangla*. Therefore, we compare the proposed approach with the other machine learning-based standard classifiers such as Naive Bayes (NB), decision tree (DT), random forest (RF), and support vector machine (SVM) to avoid bias. To implement these machine learning-based classifiers, we use scikit-learn[4] and split the entire dataset (i.e., total 4884 sentences) into 70% as the train set and 30% as the test set.

The NB classifier is trained with the prior probabilities of various classes and the likelihood of other features for every class. The training data are split recursively in the case of DT and RF. The outputs obtained from these two methods are the

[4] https://scikit-learn.org/stable/.

Table 7 Instances of sentiment predictions on BCBId dataset by our proposed approach

Bangla text	Translated English text	Normalized score	Predicted sentiment
দুজনে মনের আনন্দে খুব লিখে যাচ্ছে	Both of them are writing with great joy in their minds	0.76	Positive
হাঁদা আড়চাখে দেখছিল ভাঁদোকে প্রতিশাধে নেবার মতলবে	Handa was looking sideways to take revenge on Bhando	−0.56	Negative
এত বড় সুযাগে আর পাবেন না	You will not get this great opportunity anymore	−0.70	Negative
আমাদের পাশের বাড়িটাই ওদের	The house next to ours is theirs	0.11	Neutral

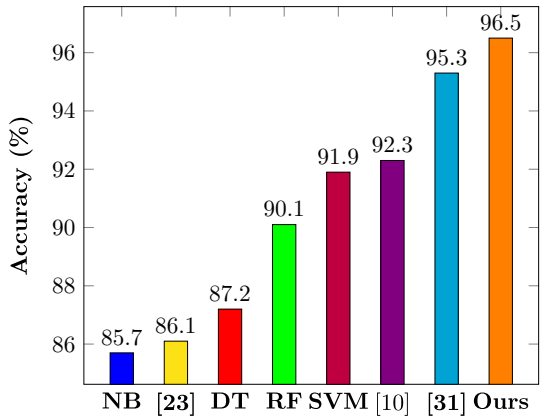

Fig. 6 Comparison based on Accuracy on BCBId

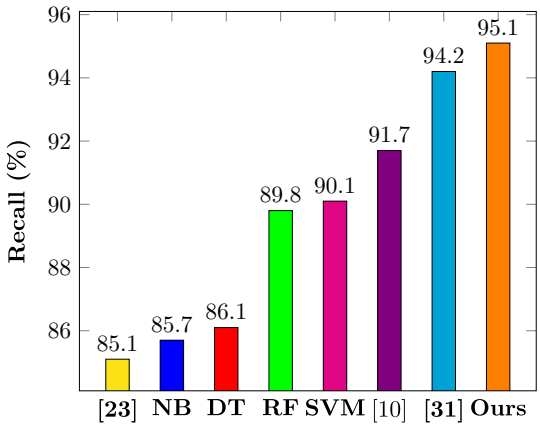

Fig. 8 Comparison based on recall on BCBId

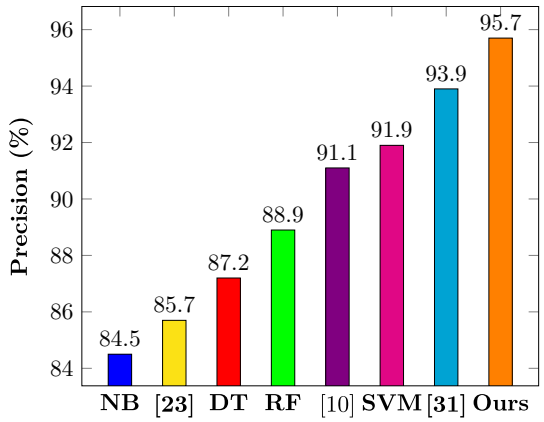

Fig. 7 Comparison based on precision on BCBId

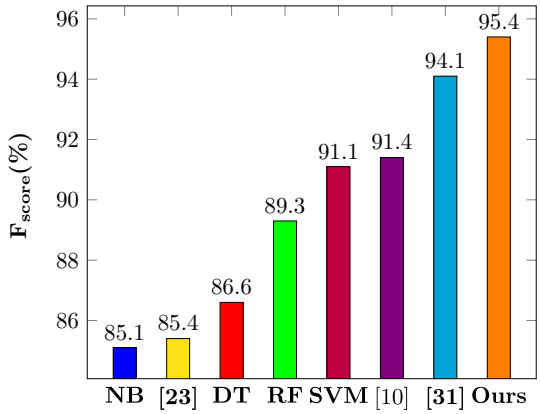

Fig. 9 Comparison based on F_{score} on BCBId

class label having the highest number of votes. In our dataset, 'linear kernels' are used in the case of SVM due to their better performance than nonlinear kernels such as polynomial or Radial Basis Function (RBF). We utilize the standard performance evaluation metrics such as accuracy, precision, recall, and F_{score}. Accuracy does not always demonstrate the complete scenarios of imbalanced data since a large number of samples create a bias toward dominant classes. So, we consider precision, recall, and F_{score} also. Besides, we also compare our method with other recently proposed lexicon-based sentiment analysis algorithms [8,22,30]. Figures 6, 7, 8 and 9 demonstrate the performance of our approach on the BCBId dataset while comparing with other methods based on accuracy, precision, recall, and F_{score}, respectively. Sentiment analysis in *Bangla* is always a challenging task due to the complex writing system, the presence of informal texts, and the wide variety of linguistic expressions. The lower number of training data, which is a constraint for comic research due to copyright issues, affects the overall performance of machine learning-based classifiers due to the

difficulties in feature learning. The proposed lexicon-based method achieves promising results in terms of all the standard metrics, i.e., accuracy, precision, recall, and F_{score}.

8 Conclusion and future work

This paper represents the first-ever collection of Bangla comic book images, the BCBId dataset, and its various applications in comic research. This dataset contains ground truth for both visual segmentation tasks and metadata encoding of every comic book page. Moreover, the description of the entire corpus has also been discussed in detail. Besides, a tool has also been designed to develop ground truth in a more comfortable and time-efficient way. Also, the applications of the dataset on various research domains of comic document image processing are explained in detail. The evaluation parameters for various tasks are also described, along with different constraints. Interested researchers can access the database by following the guidelines mentioned on the website of the BCBId Dataset for scientific research purposes only. We are continuously adding more annotations to carry out various types of research on comic books and the increment of database size.

Other than Bangla, we will also try to analyze the performance of our method of sentiment analysis on other publicly available comic datasets such as eBDtheque, Manga 109, DCM772, and COMICS.

References

1. Aizawa, K., Fujimoto, A., Otsubo, A., Ogawa, T., Matsui, Y., Tsubota, K., Ikuta, H.: Building a manga dataset "manga109" with annotations for multimedia applications. IEEE MultiMedia **27**(2), 8–18 (2020)
2. Arai, K., Tolle, H.: Method for real time text extraction of digital manga comic. Int. J. Image Process. (IJIP) **4**(6), 669–676 (2011)
3. Aramaki, Y., Matsui, Y., Yamasaki, T., Aizawa, K.: Interactive segmentation for manga using lossless thinning and coarse labeling. In: 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 293–296. IEEE (2015)
4. Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F., Iglesias, C.A.: Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst. Appl. **77**, 236–246 (2017)
5. Augereau, O., Iwata, M., Kise, K.: A survey of comics research in computer science. J. Imaging **4**(7), 87 (2018)
6. Cambria, E.: Affective computing and sentiment analysis. IEEE Intell. Syst. **31**(2), 102–107 (2016)
7. Das, A., Bandyopadhyay, S.: Sentiwordnet for indian languages. In: Proceedings of the Eighth Workshop on Asian Language Resouces, pp. 56–63 (2010)
8. Dey, A., Jenamani, M., Thakkar, J.J.: Senti-n-gram: an n-gram lexicon for sentiment analysis. Expert Syst. Appl. **103**, 92–105 (2018)
9. Digital Comic Museum. https://digitalcomicmuseum.com/. Accessed 29 May 2019
10. Dos Santos, C., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 69–78 (2014)
11. Dubray, D., Laubrock, J.: Deep cnn-based speech balloon detection and segmentation for comic books. In: ICDAR,2019, pp. 1237–1243. IEEE
12. Dutta, A., Biswas, S.: Cnn based extraction of panels/characters from bengali comic book page images. In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), vol. 1, pp. 38–43. IEEE (2019)
13. Dutta, A., Biswas, S., Das, A.K.: Cnn-based segmentation of speech balloons and narrative text boxes from comic book page images. International Journal on Document Analysis and Recognition (IJDAR) pp. 1–14 (2021)
14. Dutta, A., Zisserman, A.: The via annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276–2279 (2019)
15. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06) (2006)
16. Fukusato, T., Hirai, T., Kawamura, S., Morishima, S.: Computational cartoonist: A comic-style video summarization system for anime films. In: International Conference on Multimedia Modeling, pp. 42–50. Springer (2016)
17. Guérin, C., Rigaud, C., Mercier, A., Ammar-Boudjelal, F., Bertet, K., Bouju, A., Burie, J.C., Louis, G., Ogier, J.M., Revel, A.: eBDtheque: a representative database of comics. In: ICDAR, pp. 1145–1149. IEEE (2013)
18. Gupta, V., Detani, V., Khokar, V., Chattopadhyay, C.: C2vnet: A deep learning framework towards comic strip to audio-visual scene synthesis. In: International Conference on Document Analysis and Recognition, pp. 160–175. Springer (2021)
19. Hartel, R., Dunst, A.: An ocr pipeline and semantic text analysis for comics. In: International Conference on Pattern Recognition, pp. 213–222. Springer (2021)
20. He, Z., Zhou, Y., Wang, Y., Wang, S., Lu, X., Tang, Z., Cai, L.: An end-to-end quadrilateral regression network for comic panel extraction. In: Proceedings of the 26th ACM international conference on Multimedia, pp. 887–895 (2018)
21. Ho, A.K.N., Burie, J.C., Ogier, J.M.: Panel and speech balloon extraction from comic books. In: 2012 10th IAPR international workshop on document analysis systems, pp. 424–428. IEEE (2012)
22. Hossen, M., Dev, N.R., et al.: An improved lexicon based model for efficient sentiment analysis on movie review data. Wirel. Pers. Commun. **120**(1), 535–544 (2021)
23. Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daume, H., Davis, L.S.: The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7186–7195 (2017)
24. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. J. Artif. Intell. Res. **50**, 723–762 (2014)
25. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of Massive Data Sets. Cambridge University Press (2020)
26. Li, L., Goh, T.T., Jin, D.: How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. Neural Comput. Appl. **32**(9), 4387–4415 (2020)
27. Li, L., Wang, Y., Gao, L., Tang, Z., Suen, C.Y.: Comic2cebx: A system for automatic comic content adaptation. In: IEEE/ACM Joint Conference on Digital Libraries, pp. 299–308. IEEE (2014)
28. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimed. Tools Appl. **76**, 21811–21838 (2017)

29. Matsui, Y., Yamasaki, T., Aizawa, K.: Interactive manga retargeting. In: ACM SIGGRAPH 2011 Posters, pp. 1–1 (2011)
30. Mowlaei, M.E., Abadeh, M.S., Keshavarz, H.: Aspect-based sentiment analysis using adaptive aspect-based lexicons. Expert Syst. Appl. **148**, 113234 (2020)
31. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Sentiful: a lexicon for sentiment analysis. IEEE Trans. Affect. Comput. **2**(1), 22–36 (2011)
32. Nguyen, N.V., Rigaud, C., Burie, J.C.: Comic characters detection using deep learning. In: ICDAR ,2017, vol. 3, pp. 41–46. IEEE
33. Nguyen, N.V., Rigaud, C., Burie, J.C.: Digital comics image indexing based on deep learning. J. Imaging **4**(7), 89 (2018)
34. Nguyen, N.V., Rigaud, C., Burie, J.C.: Comic MTL: optimized multi-task learning for comic book image analysis. Int. J. Document Anal. Recogn. (IJDAR) **22**(3), 265–284 (2019)
35. Nguyen, N.V., Vu, X.S., Rigaud, C., Jiang, L., Burie, J.C.: Icdar 2021 competition on multimodal emotion recognition on comics scenes. In: ICDAR,2021, pp. 767–782. Springer
36. Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T., Aizawa, K.: Object detection for comics using manga109 annotations. Preprint arXiv:1803.08670 (2018)
37. Pang, X., Cao, Y., Lau, R.W., Chan, A.B.: A robust panel extraction method for manga. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 1125–1128. ACM (2014)
38. Qian, Q., Huang, M., Lei, J., Zhu, X.: Linguistically regularized lstm for sentiment classification. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1679–1689 (2017)
39. Qin, X., Zhou, Y., He, Z., Wang, Y., Tang, Z.: A faster r-cnn based method for comic characters face detection. In: ICDAR, vol. 1, pp. 1074–1080. IEEE (2017)
40. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR, pp. 779–788 (2016)
41. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: CVPR, pp. 7263–7271 (2017)
42. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: NIPS, pp. 91–99 (2015)
43. Rigaud, C., Burie, J.C., Ogier, J.M.: Text-independent speech balloon segmentation for comics and manga. In: International Workshop on Graphics Recognition, pp. 133–147. Springer (2015)
44. Rigaud, C., Burie, J.C., Ogier, J.M., Karatzas, D., Van de Weijer, J.: An active contour model for speech balloon detection in comics. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 1240–1244. IEEE (2013)
45. Rigaud, C., Guérin, C., Karatzas, D., Burie, J.C., Ogier, J.M.: Knowledge-driven understanding of images in comic books. IJDAR **18**(3), 199–221 (2015)
46. Rigaud, C., Le Thanh, N., Burie, J.C., Ogier, J.M., Iwata, M., Imazu, E., Kise, K.: Speech balloon and speaker association for comics and manga understanding. In: ICDAR,2015, pp. 351–355. IEEE
47. Rigaud, C., Tsopze, N., Burie, J.C., Ogier, J.M.: Robust frame and text extraction from comic books. In: International Workshop on Graphics Recognition, pp. 129–138. Springer (2011)
48. Sun, W., Burie, J.C., Ogier, J.M., Kise, K.: Specific comic character detection using local feature matching. In: ICDAR, 2013, pp. 275–279. IEEE
49. VGG image annotator. http://www.robots.ox.ac.uk/~vgg/software/via/via.html. Accessed 11 March 2019
50. Walsh, J.A.: Comic book markup language: an introduction and rationale. Digital Humanities Q. **6**(1) (2012)
51. Wang, Y., Zhou, Y., Tang, Z.: Comic frame extraction via line segments combination. In: ICDAR,2015, pp. 856–860. IEEE
52. Xie, M., Xia, M., Liu, X., Wong, T.T.: Screentone-preserved manga retargeting. Preprint arXiv:2203.03396 (2022)
53. Yadav, A., Vishwakarma, D.K.: Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev. **53**(6), 4335–4385 (2020)