WILEY | Hindawi

Research Article

The Analysis of Public Opinion in Colleges and Universities Oriented to Wireless Networks under the Application of Intelligent Data Mining

WenNing Wu [1] **and ZhengHong Deng** [2]

[1]*International Cooperation Department/School of Cybersecurity, Northwestern Polytechnical University, Xi'an, 710072 Shaanxi, China*
[2]*School of Automation/ School of Cybersecurity, Northwestern Polytechnical University, Xi'an, 710072 Shaanxi, China*

Correspondence should be addressed to WenNing Wu; wuwenning@nwpu.edu.cn

In recent years, the incidence of public opinion in colleges and universities has been high. Monitoring, forecasting, and responding to public opinion of students in colleges and universities have increasingly become the work that education management departments at all levels attach great importance to. For each university, how to understand the sensation of teachers and students in real time in the era of informationization entering the intelligent campus has become an urgent problem. How to collect college campus network information, analyze and manage this information, and find hot topics from it has a profound impact on the reform of colleges and universities. Hence, in this paper, we propose a public opinion analysis framework based on intelligent data mining technique. Its advantage lies in the fact that it can withdraw the needed and unknown knowledge and regularities from the massive network data and host log data. It is a new attempt to use data mining in achieving public opinion. At present, data mining algorithm applied to public opinion analysis mainly has four basic patterns: association, sequence, classification, and clustering. Data mining technology is advanced for: it can process large amount of data. It does not need the users' subjective evaluation and is more likely to discover the ignored and hidden information. Here, initially, the dataset is collected, which is preprocessed and divided into a training set and test set. Feature extraction of the text is done using Linear Discriminant Analysis (LDA). After that, text cosine similarity calculation is performed to compute the similarity between text vectors obtained from the LDA. Convolutional neural network (CNN) is used for classification purpose. We proposed Krill Herd Harmony Search Optimization Algorithm (KHHSOA) for optimizing the CNN and classifying the text into positive and negative opinion. The proposed system is simulated using MATLAB simulation tool, and the performance is analyzed in terms of metrics like accuracy, precision, recall, F-measure, kappa coefficient, and error rate. The proposed method is proved to be better when compared with the existing techniques.

1. Introduction

It is critical for college students to have a mature view on life and morals throughout their early years in college. In college, students' perceptions of life and other factors, including public opinion, might impact their ideology. When it comes to media, the Internet has made it easier for consumers to get the information they want at any time and from any location [1]. However, the quality and dependability of information are becoming more and more pronounced.

'Students' mental health may be negatively impacted by the spread of false and bad information in the college community. Students who are hooked to the Internet and wireless mobile devices, such as smart phones, have been shown to have higher levels of tension and anxiety and worse levels of scholastic achievement and overall happiness, according to personality study [2]. Having a negative outlook and feeling that your life is useless may lead to depression and even suicide in certain cases. Social media and the government should be seen as liars by students because of the widespread

dissemination of misinformation. Students' suspicion of the government will lead to a lack of harmony when they join the workforce after graduation. There is considerable evidence that students who thrive in a pleasant atmosphere are more likely to do great things in college, and this is based on a student's mental state. A college's role in preventing students from being negatively affected by the explosion of information is to focus on providing support and guidance to those students who are struggling, to take responsibility for helping them correct their outlook on life and their values, and to inspire students to fight for the advancement of humanity as a whole [3]. Students, on the other hand, are often reluctant to seek help because they do not want to be seen as "sick." This necessitates that universities take a more active role in guiding pupils. If students try to conceal their issues, it will be difficult for universities to recognise who needs to be aided when they face thousands of students. The convolutional neural network (CNN) is a deep learning (DL) technology that may be used to predict the mood of children. DL tools can quickly and accurately predict students' state of mind if they have enough training data to do so [4]. Colleges may tailor their counseling to students' individual characteristics if they are able to accurately forecast their mental state. Predicting people's opinions on things by data analysis acquired from the Internet has been extensively utilized; however, it may not be as effective for predicting youths' mental states. The reason for this is because it is difficult to accurately forecast a student's emotional condition:

(1) Subject focus: this work is focused on the people who make judgments, but not the judgments they have made

(2) Multiaspect: to improve the accuracy of the analysis, the predictor should acquire a variety of facts, but students should not post it online

(3) Low consistency on aspects: because various pupils have varied interests, it is unwise to establish a public criteria of a certain solution. A single sample should encompass most areas of a student's everyday life, and the data from multiple samples should be consistent in substance to fulfill these criteria. The improved CNN technique is the data mining technology used as a forecaster. A predictor's generalization performance, or prediction accuracy on new samples, is one of the most essential criteria. Samples are too small in size to fit the sample area. Lack of samples may lead to overfitting in CNN training. On new samples, an overfitted CNN performs poorly. So a new approach is required. This study will provide a method for improving the testing process using the Krill Herd Harmony Search Optimization Algorithm. Section 2 reviews the related work. Section 3 introduces the problem statement. Section 4 describes the process of using optimization to improve the CNN performance over opinion mining. Section 5 describes the details of the experiment and shows the results. Section 6 concludes the paper.

2. Literature Survey

In reference [1], the COVID-19 job promotion policy for college graduates is the study object, and the PMC index model is used to filter the policy texts, resulting in two ideal policy texts, and Weibo comments are used to develop the evaluation model of policy measure support degree. The findings reveal that the public's level of support for COVID-19's job promotion programs for college graduates needs to increase. With regard to position and transfer measures, public opinion seems to be neutral, while it is clearly displeased with subsidy and channel initiatives [2]. Analyze the effect of online public opinion on the image and reputation development of colleges or universities using simple statistical methods without taking into account other factors, such as the hotness characteristics of online public opinions and semantic information. The purpose of this research is to enhance the analysis performance of online public views in colleges and universities utilizing short-term trend prediction findings based on the Public Opinion Hotness Index and time series-based trend analysis [3]. In the 2019 coronavirus illness thoughts and feelings of Koreans and Japanese (COVID-19), it looked at the frequency of certain terms in tweets on COVID-19, as well as changes in the interests of Korean and Japanese users [4]. There are five primary areas of school affair information that will be analyzed in this study: student, teaching, research, school affairs, and financial affairs. Data exploration will be used to examine this information [5]. A study on fuzzy sets in public opinion analysis from the viewpoint of big data explores the diverse functions of fuzzy sets in network public opinion analysis from four aspects: broad fuzzy collection of fuzzy terms, fuzzy reasoning, fuzzy granularity, and fuzziness as a whole [6]. Tweets from three prominent coronavirus-related hashtags (#COVID19, #Coronavirus, and #SARSCoV2) are used to examine public opinion and attitudes concerning the COVID-19 outbreak [7]. Here, an investigation of the year-to-year patterns of relevant articles from the Web of Science (WoS) database is the goal of this study. Several institutions from China, the United States, Spain, and South Korea have made a significant contribution, according to the results. Interdisciplinary cooperation is possible because of research's ability to create a bibliographic linking network of publications [8]. A large amount of Twitter text relating to certain items and services is evaluated, and the general opinions of the users of those objects are summarised to assist producers and servers in their tasks [9]. The features of the various stages of research investigations are outlined. These summaries provide researchers excellent tools for opinion mining and evaluation criteria. Questionnaires have been extensively utilized to gather data in various circumstances when it is required to develop a person's whole personality profile [10], combining opinion mining and sentiment analysis to create a teacher performance assessment tool. Based on the positive and negative comments from students in either English or Filipino, the research may be able to determine the strengths and shortcomings of faculty members. Sentiment score and numerical response rating for teachers' evaluations are provided by a suggested

system that analyzes qualitative and quantitative data. Additionally, it will show the proportion of good and negative comments given by the pupils [11] and examines student responses from both course surveys and Internet sources for sentiment polarity, the emotions conveyed, and satisfaction or disapproval levels. The system's dependability may be shown by comparing its findings to those of direct assessments [12], provides a mechanism for computing students' emotion polarity from their remarks, and introduces readers to OM. A toy example is provided to demonstrate how the idea works in reality once the logic and mathematical formalization have been explained [13]. Here, for the categorization of educational views in an Intelligent Learning Environment (ILE-Java), this paper compares sentiment analysis classifiers based on machine learning, deep learning, and an evolutionary method dubbed EvoMSA. Students' feelings towards professors, examinations, homework, and academic projects, among other things, are reflected in two corpora of phrases in the programming languages area. sentiTEXT and eduSERE both have polarity labels (positive and negative), but the eduSERE corpus contains positive and negative learning-centered emotions (engaged, enthusiastic, bored, and annoyed) [14]. Student feedback datasets often include context patterns that may be predicted using an effective fusion-based neural network (EF-NN) classifier. A support vector machine and a convolutional neural network are combined in our new EF-NN model. According to attribute factors such as the interaction between students, examinations, and notes offered, a student feedback dataset is retrieved [15]. To make collecting feedback from students easier and more efficient, a student opinion mining system has been developed. This app's importance lies in the fact that it is less priced and offers a more private means of gathering student feedback. Python, Scikit-Learn, Textblob, Pandas, and SQLite are among the most important technologies used in the creation of this application. Django has a built-in server that enables the app to execute locally instead of across the network [16]. Using the concepts of ensemble learning and deep learning, we develop an effective sentiment classification technique for MOOC evaluations with strong predictive performance. Many research topics on sentiment analysis of educational data are addressed in this paper. Conventional supervised learning, ensemble learning, and deep neural networks have all been tested for their predictive abilities. For sentiment analysis on MOOC assessments, the efficacy of text representation systems and word-embedding techniques has also been examined [17]. The idea was to create a system that could automatically classify students' input. Higher education institutions benefit greatly from this method since it conserves resources such as time, people, and money. After two years of data collection from university students, they categorized the data into three categories: positive, negative, and neutral. They used 5000 categorized sentences to create the Vietnamese sentiment dataset. Our annotated data is next analyzed using three classifiers: Naive Bayes (NB), Maximum Entropy (ME), and support vector machine (SVM) [18]. Introduce a free text review mechanism for ABSA in the Serbian language for student opinion polls. The tiniest text granularity—the

level of sentence fragment—was used for sentiment analysis [19]. The search was guided by a PRISMA framework that looked for papers published between 2015 and 2020 in scientific literature electronic databases. They selected 92 papers out of 612 on the sentiment analysis of students' comments in learning platform settings that were relevant to their research. Even though obstacles have been recognized, this subject is quickly expanding, particularly with relation to the use of DL, which is the most current trend. Many factors need to be addressed in order for research and development in the sector to become more mature. In addition, they emphasized the need of organized datasets, standardized solutions, and enhanced attention to emotional expression and detection [20]. Sentiment analysis based on machine learning is shown here. They used traditional text representation approaches and machine learning classifiers to assess a 700-student review corpus written in Turkish [21]. The classic TF-IDF approach is enhanced by adding the characteristic word's position weight (span weight) [22]. Examine sequential deep neural networks' function in multimodal sentiment analysis. For textual, visual, and multimodal SA employing RNN and architectural variations, we give a comprehensive overview of the problems, concerns, and techniques [23]. Here, by analyzing the subject distribution of postings on the Baidu Post Bar (BPB), they initially crawled data from 10 universities' post bars and used the LDA algorithm to find topics. They defined word weight by TF-IDF, length, cover, and title. On the other hand, they established the subject heat ranking methodology based on support papers, replies, and time. The themes' meaning was represented by two label words then studied subject distributions [24]. During the COVID-19 shutdown, a study examines the potential of distant learning to save the academic year for students. It examines the use of several online application tools during the coronavirus infectious disease 2019 (COVID-19) pandemic to evaluate the adoption of distant learning. As the COVID-19 virus ravages major cities throughout the world, distance learning flourishes as a valid alternative to classroom teaching. Distance learning systems employing online resources have been implemented by educational institutions in an effort to save the academic year. It was shown that confidence in official and informal information sources, awareness of remote learning, and willingness to accept distance learning were all factors in this research. When it comes to remote learning, they propose that situational awareness and preparedness are obtained via confidence in information sources connected to distant learning [25]. New insights into the feedback process may be gained by analyzing PA data using learning analytics. As part of this exploratory project, we will be looking into backward evaluation, which is an underresearched facet of peer assessment. Backward evaluation qualities and the link between rubric characteristics and backward evaluation are examined. Statistical and epistemic network analyses were used to examine a large dataset ($N = 7,660$ entries) from an online platform called Peergrade. By acknowledging their mistakes, praising its usefulness, and expressing their confusion about its meaning, students

who found feedback useful were more accepting, whereas students who found feedback less useful were more defensive, saying they were unsure of its meaning, critical of its form, and disagreed with the claims.

3. Problem Statement

The Web is an excellent platform for online learning and sharing thoughts, viewpoints, and ideas for organizations or goods because of the abundant bandwidth accessible through wireless. Internet discussions are vital while choosing a business or product. Computational analysis of online evaluations, comments, attitudes, and opinions has become obligatory. It requires some thinking to upgrade and evaluate the user's different viewpoints that are conveniently available on the sites. Opinion mining studies user views of goods, services, and the learning system. This reveals the outlook's optimism or pessimism. The basic purpose of opinion mining is to identify the polarity of feedbacks by deducing object attributes and components. Growth of mobile communication and tiny devices has resulted in an innovative and distinctive teaching style. The deep learning system's learning conditions may be accessed at any time to assist learners enhance their learning abilities. As a consequence, while designing a deep learning system, consider the user's viewpoint. This involves employing self-activating text evaluation to produce views and determining the mood of those views from the perspective of people inspecting or assessing services. It has several restrictions, limited categorization reliability, and no document-level views. This proposed research study addresses all of these issues.

4. Proposed Methodology

Most college and university students utilize social media to communicate their emotions and opinions in their daily life as a group with active thoughts. The massive expansion in the use of social media has resulted in a surge in public opinion. Teachers would benefit from analyzing public opinion from social media in order to gain a better understanding of their students' ideas and be advised for ideological teaching. For an efficient investigation of public opinion, this paper employs an intelligent data mining technique. Figure 1 depicts a schematic illustration of the proposed methodology.

4.1. Dataset Collection. This paper focuses on obtaining data from the Baidu Post Bar (BPB) college post dataset. Baidu Post Bar is China's largest network community, bringing together people who share the same interest. The majority of college students enjoy posting their thoughts on their college or university's message boards. Baidu, the Chinese online service firm, hosts the most popular Chinese communication platform. On December 3, 2003, Baidu Tieba was launched as a subsidiary of Google. An online community that mainly relies on Baidu's search engine has been created. Users may use the "bar" search function to find a forum on a certain subject, and if no such forum exists, one will be established. By 2015, Baidu Tieba had over 300 million monthly active users and 1.5 billion registered users. Baidu Tieba has 23,254,173 communities as of June 6, 2021. According to the "Ministry of Education Sunshine Entrance Open Platform," China had 2,631 colleges and universities as of May 31, 2017. The majority of colleges have a post bar that students attend. The post bar's dataset is quite vast. In the Zhengzhou University Bar, for example, there are around 800,000 themes and twenty million posts. It is hard to access post data from all two thousand universities because of our restricted computer capacity. As a result, we only chose the top 10 institutions having the most members on BPB, as shown in Table 1.

Using a crawler, we can locate the post bar websites of the universities listed in Table 1. We crawled the data from every post bar website we could find. The title, content, post time, poster, number of replies, and reply time are all included in each post. For these colleges, the total number of positions is around 90,000.

4.2. Preprocessing of Data. Data preparation is the most basic and important step in data mining. Because there is no single data source for the original data and the format is not standard, noise data, as well as the presence of erroneous data, will have a significant impact on the classification result. Invalid and inaccurate data are constantly encountered during the crawling process; as a result, these data are rarely used in the analysis process or create significant problems in the results, and they must be preprocessed. Repeatedly gathered data, repeated expressive words, shorter sentences, nonsensical sentences, or ambiguous sentences are mostly deleted during data preprocessing.

Filtering noisy information and deleting stop words for public opinion data characteristics are the two processes of data preprocessing.

4.2.1. Filtering Noise Information. There is relatively little information in BPB that is overly complicated, has a high repetition rate, and is actually practical, due to the grassroots, autonomy, and repeated nature of text material. As a result, how to filter out the worthless data in BPB will be critical in future research. The main objective of filtering the noise information is to exclude filtering out some information that may be useless for the research by mining the original data and picking the appropriate processing procedure according to their own unique format. The noise filtering of data is usually classified into two groups for the preprocessing procedures of public opinion material. Its purpose is to filter the text content of the public opinion as well as some other information outside of the public opinion content, such as the BPB user.

4.2.2. Removing Stop Words. Stop words are words that have no differentiation or ability to describe (such as "bar," "and," and "have"). These words have no practical value for text processing by definition. These words are largely virtual words that exist with identical probability in various sorts
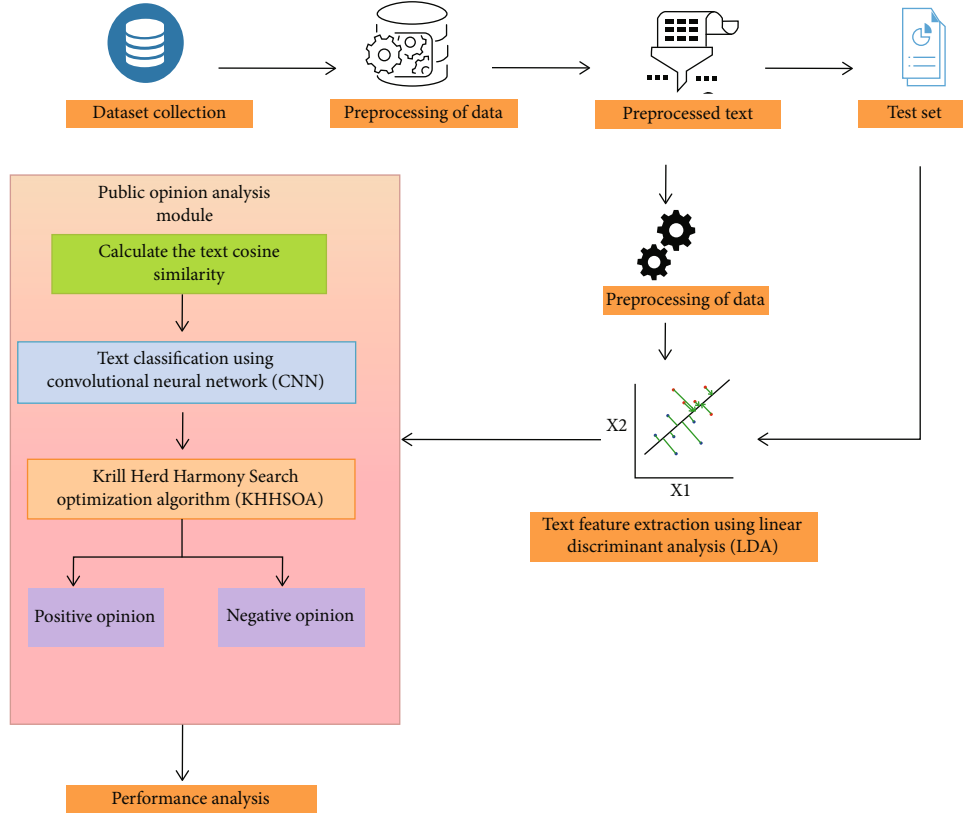
FIGURE 1: Schematic representation of the proposed method.

TABLE 1: Dataset description.

Colleges & universities	Members' number
Sichuan conservatory of music	137820
Zhengzhou University	90587
Sias International University	78549
Southwest Jiaotong University	60883
Henan University	60590
Sichuan University	53418
Wuhan University	51321
Henan University	45262
Xiamen University	45094
Central South University	44563

of texts, and they also increase the dimension of the text feature vector; thus, they must be deleted.

The training set and test set are created first when the experimental data is obtained. The training set is mostly utilized in machine learning, and it generates characteristic data and multituples, as well as constructing the bag of words for emotion analysis.

4.3. Text Feature Extraction Using Linear Discriminant Analysis (LDA).
The feature extraction model in this paper is the Linear Discriminant Analysis (LDA) model. As stated in Equations (1) and (2), the LDA model characterizes public opinion as a sequence of semantically connected terms and their distributions on the opinion:

$$Q = \{(X_i, p(X_i \mid q, \alpha)) \mid 0 < i < M\}, \tag{1}$$

$$p(X_i \mid q, \alpha) \sim \text{multinomial}(\alpha), \tag{2}$$

where Q is an opinion; X_i is the ith word in the vocabulary; M is the size of the vocabulary; and $p(X_i \mid q, \alpha)$ is the probability of the ith word in opinion Q. The larger the $p(X_i \mid q, \alpha)$ is, the more relevant X_i and q are. The word distribution is multinomial with the parameter α.

The LDA model-defined document is a sequence of opinions and their distributions in the document, shown as Equations (3) and (4):

$$s = \left\{ \left(q_j, p\left(q_j \mid s \right) \right) \mid 0 < j < N \right\}, \tag{3}$$

$$q_j \sim \text{multinomial}(\theta), \tag{4}$$

where s is the document; q_j is the jth opinion in the opinion space; N is the opinion number; and $p(q_j \mid s)$ is the probability of the jth opinion in a document s. The larger the $p(q_j \mid s)$ is, the more relevant the opinion j and document s are. The opinion distribution is also a multinomial distribution with a Dirichlet random variable, as shown in the following equation:

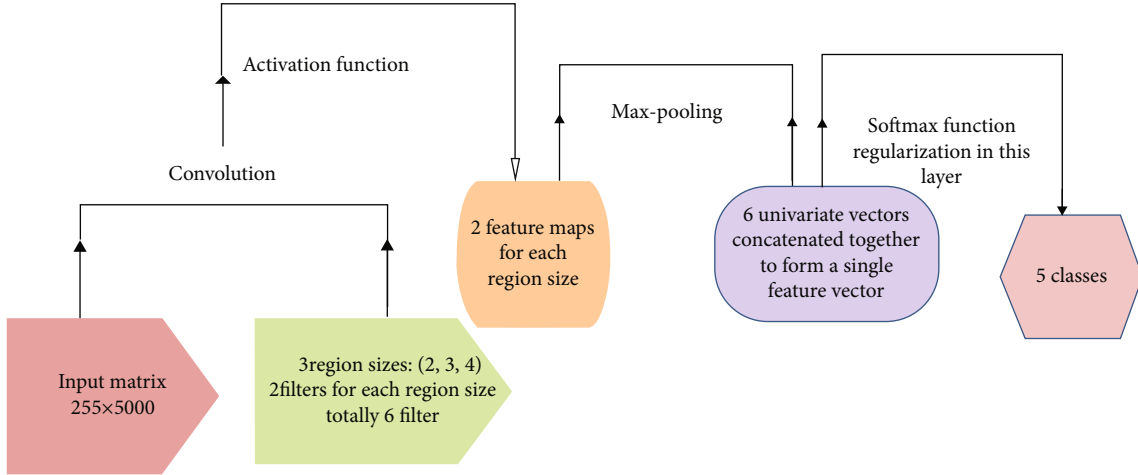$$\theta \sim \text{Dir}(\delta). \tag{5}$$

FIGURE 2: Illustration of a convolutional neural network (CNN) architecture for text classification.

Given the parameters a and b, the opinion distribution for the jth document is θ_j:

$$\theta_j = \left(\theta_{j1}, \theta_{j2}, \cdots, \theta_{jN}\right). \tag{6}$$

The word distribution for j^{th} topic is δ_i:

$$\delta_i = \left(\delta_{i1}, \delta_{i2}, \cdots, \delta_{iM}\right). \tag{7}$$

Gibbs sampling can be used to solve the parameters θ and δ:

$$\begin{cases} \theta_{jn} = \dfrac{\text{count}_j^n + \sigma_n}{\sum_{n=1}^{N} \text{count}_j^n + \sigma_n}, \\[3mm] \delta_{mn} = \dfrac{\text{count}_m^n + \gamma_n}{\sum_{m=1}^{M} \text{count}_m^n + \gamma_n}, \end{cases} \tag{8}$$

where count_j^n is the number of words that are included in the opinion n of the jth document and count_m^n is the number of times that word n occurs in opinion m.

Before the inference parameters are determined, we should set the topic number M.

This paper uses perplexity to estimate M:

$$\text{Perplexity} = \exp\left(\frac{\sum_{j=1}^{S} \log\left(p(s_j)\right)}{\sum_{j=1}^{S} L_j}\right), \tag{9}$$

where S is the document number, L_j is the length of the jth document, and $p(s_j)$ is the probability of generating the jth document. The smaller the perplexity is, the better the results are.

The super parameters a and b can be defined by the opinion number and vocabulary size as Equations (10) and (11):

$$\gamma = \frac{50}{R}, \tag{10}$$

$$\delta = \frac{200}{S}. \tag{11}$$

4.4. Text Cosine Similarity Calculation. In general, the keywords for the same topic network public opinion are about the same, while the keywords for the different topic network public opinions are not the same. As a result, the similarity of public opinion on the same topic network is higher, while the similarity of public opinion on other topic networks is lower. It can calculate similarity and develop a similarity measure model of public opinion based on this feature. At the same time, high similarity of network public opinion is regarded as the same theme network public opinion, while low similarity is judged as a separate topic network public opinion, according to the principle of cluster analysis. As a result, calculating similarity is critical for network public opinion. In document comprehension tasks, similarity measurement is a significant computing overhead, and cosine similarity is one of the most widely used text similarity methods. We use Equation (12) to calculate log frequency weight of term t in opinion d.

$$v_{p,b} = \begin{cases} 1 + \log_{10} pf_p, b, & \text{if } pf_p, b > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

When two papers are almost identical, cosine similarity returns one, and when the documents are extremely distinct, it returns zero. We must normalize two papers x and y to one in L2 norm in order to find cosine similarity between them.

$$\sum_{j=1}^{n} y_j^2 = 1. \tag{13}$$

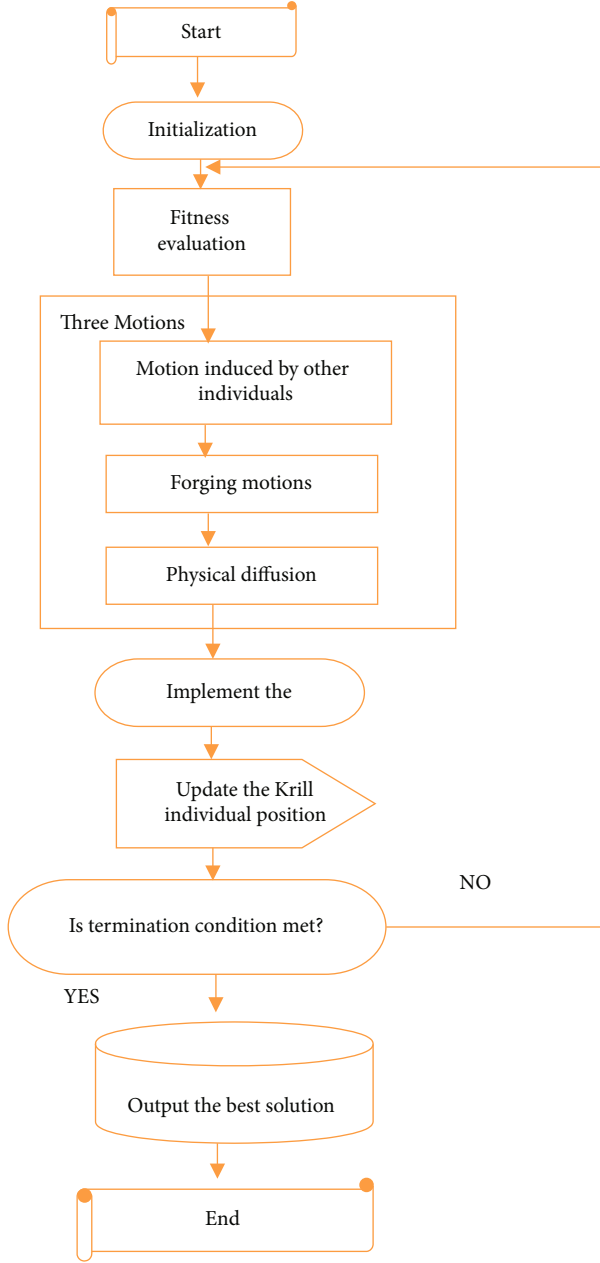By having two normalized vectors x and y, the cosine similarity between them will be simply the dot product of them.

Figure 3: Flow chart of the KHHSOA.

$$\cos(x, y) = \frac{\sum_{j=1}^{n} x_j y_j}{\sqrt{\sum_{j=1}^{n} x_j^2} \sqrt{\sum_{j=1}^{n} y_j^2}}. \tag{14}$$

Careful examination of Equation (14) shows that cosine similarity is directly derived from Euclidean distance.

$$d_{\text{Euclid}}(x, y) = \left[\sum_j \left(x_j - y_j \right)^2 \right] = \left[2 - 2 \sum_j x_j y_j \right]. \tag{15}$$

To find out the cosine similarity between the two planes, the Euclidean distance was calculated by using the above equation (15).

Figure 4: Number of comments vs. error rate (negative opinion).

4.5. Text Classification Using Convolutional Neural Network (CNN). After the pooling operation, a full connection layer is used to classify the results of the convolutional neural network, which is a nonlinear activation function applied to the outcomes of the convolutional operation. The filter, often known as the kernel function, is at the heart of convolutional operations. By sliding from top to bottom and left to right in the original matrix, it completes feature extraction. The width of the kernel function in public opinion mining is typically equal to the width of the original matrix, and the kernel function only slides in the upper and lower directions, ensuring the word's integrity as the smallest granularity in the opinion.

The usual convolutional layer, pooling layer, and fully connected layer are incorporated in our proposed CNN model, as shown in Figure 2.

On 255 words, we utilized a one-layer CNN with a word vector of dimension 111,728.

(1) Input matrix: because 255 is the maximum number of stems included in each document in the corpus after the preprocessing stage, we fixed the input in 255 words. The word vectors have a dimension of 111,728, which corresponds to the total number of documents in the corpus. We now have an input matrix of the shape $s \times d$ or $255 \times 111{,}728$ with s denoting the number of words and d denoting the dimension of the word vector

(2) Filters: one of CNN's desired qualities is that it maintains orientation, which is beneficial to us because texts have a one-dimensional structure in which the order of words is important. To match the word vectors, one dimension of the filter was fixed for all filters, while the region size "h" was adjusted from one filter to the next. The number of rows in the input matrix that represent words is referred to as
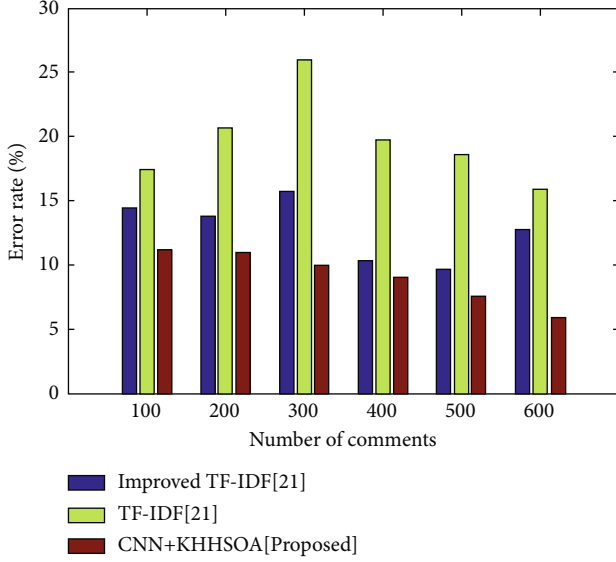
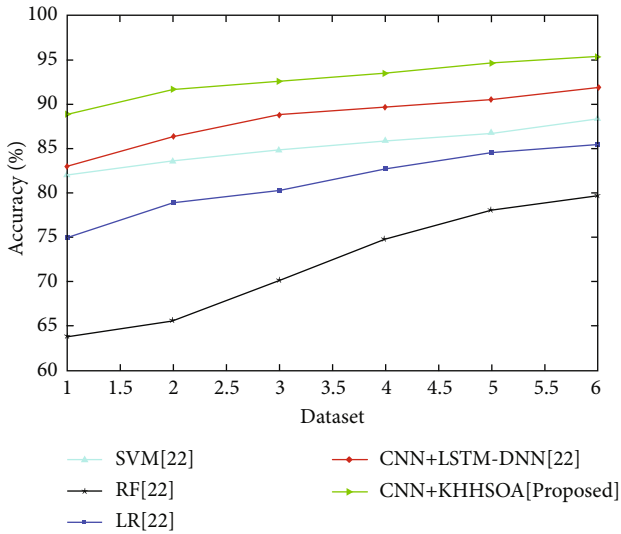Figure 5: Number of comments vs. error rate (negative opinion).

Figure 6: Dataset vs. accuracy.

Figure 7: Dataset vs. *F*-measure.

Figure 8: Dataset vs. kappa coefficient.

region size. The illustrations of the filters in Figure 2 are not the outcomes of filtering the input matrix. We choose to use three different area sizes here, with each recovering two, three, or four words (stems) at a time. Furthermore, two filters were chosen for each region. In general, there are six filters

(3) Feature maps: in this section, we will go over how the CNN performs convolutions and filtering. We have loaded the input matrix with some pretreatment findings, and we have filtered the filter matrix with certain numbers to make the explanation clearer

(4) Max-pooling: the dimensionality of c is dependent on both s and h; i.e., it will change as a result of filters
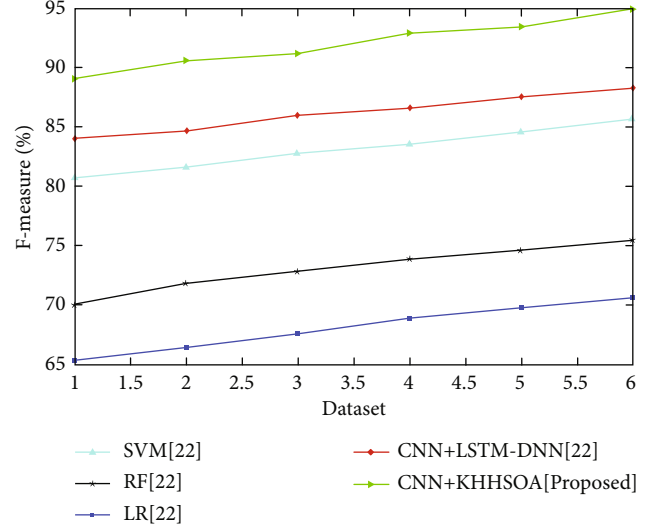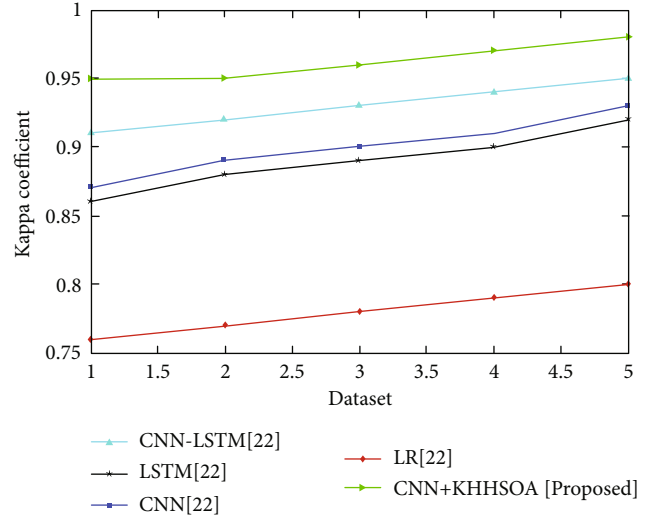
with different region sizes. We utilized the max-pooling function to solve this problem, extracting the largest number from each c vector

(5) Softmax: the resulting vector has a fixed length of 6 entries after max-pooling (number of filters). To execute the classification task, this resulting vector can be input into a fully connected (softmax) layer

4.6. Krill Herd Harmony Search Optimization Algorithm (KHHSOA). The Harmony Search (HS) algorithm is a population-based stochastic search technique, similar to the Krill Herd (KH) algorithm. The algorithm keeps track of a set of solutions, each of which indicates a potential solution to the problem. Therefore, for the data classification problem, we suggest the Krill Herd Harmony Search Optimization Algorithm (KHHSOA), which is a mix of the Krill Herd algorithm and the harmony search algorithm.
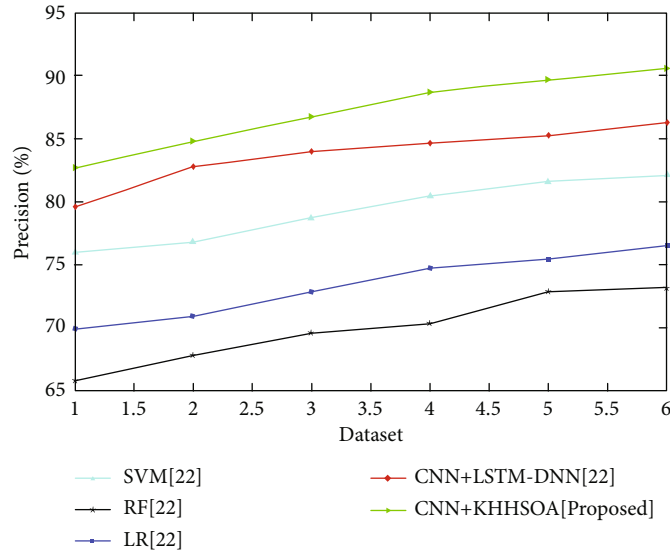
FIGURE 9: Dataset vs. precision.

The KHHSOA periodically creates a population of size $S * n$. In the event of the Krill Herd (KH) algorithm, such individuals are viewed as a herd, while in the event of the Harmony Search (HS) method, they are considered as harmonies. KH actions feed krill members into Krill Herd Memory (KHM) to construct memory of size $S * n$. The krill members are then classified according to their fitness value and given into Harmony Search in order to improve KHHSOA by developing a new solution and introducing it to the population if it is superior to the worst solution in KHM. As a result, KHHSOA may use superior personnel in every iteration to find the best answer and keep improving it. Furthermore, krills with low performance continue to use the HS regulation to prevent early integration. Figure 3 depicts the architecture of the KHHSOA.

5. Result and Discussion

This section illustrates the outcome of opinion analysis. The method was performed using the resource for emotional analyses—the natural language—which distinguishes positive and negative opinions. The proposed model achieves good dataset performance, which evaluates the efficiency of the pretrained model. To calculate the experimental effects, (a) precision, (b) recall, (c) F1-score, (d) accuracy, and (e) kappa coefficient are used to evaluate the output.

Such performance measures were taken to determine the opinion classification outcomes. While focused on essential results, the proposals were correlated with other algorithms and methods of classification. Because of the greater effectiveness of this optimized CNN, the highest result can be obtained by utilizing the proposed optimized classifier to identify positive and negative opinions than all other current approaches and forecast the term output. The findings suggest that the new approach has increased its efficiency. For 96.5% recall, a cumulative precision of 91% was calculated along with the 99.8% kappa score. The *F*-measure will then increase from 95%.

Nouns and adjectives express the sentence's emotions. The proposed approach lends greater weight to nouns and adjectives in phrases, making the retrieved distinctive terms more suitable for further opinion analysis (Figures 4 and 5). Also, the quantity of comments on the test set increases the emotional two-level marked precision. Because each word completion indicated the opinion level, the recommended approach has a low error rate over opinion analysis. To prove the efficiency of the suggested classifier, it can be compared with the existing mechanisms [21, 22] over some performance metrics.

5.1. Accuracy. Accuracy determines the errors ordered; compute the arithmetical field of the composition. Low precision causes a difference between a calculation and a "real" value. The exceptional data samples are tested several times by the same algorithm, and a correct test is performed on the machine or system. The precision of the final information is the proportion of the actual data.

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{FalseNegative} + \text{FalsePositive}}.$$
$$+\text{TrueNegative}$$

$$(16)$$

5.2. Precision. The precision can be calculated by using the following equation:

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}. \quad (17)$$

5.3. Recall. The recall can be calculated by using the following equation:

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}. \quad (18)$$
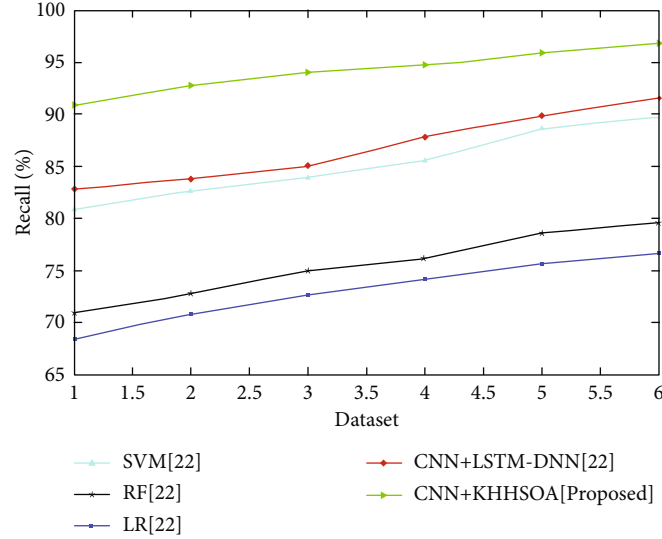
FIGURE 10: Dataset vs. recall.

5.4. F-Score. The accurate measurement of the method by believing both precision and recall factors is referred to as F1-score

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{19}$$

5.5. Kappa Statistic. The kappa statistics are an important parameter for assessing the model accuracy. The outcome of the suggested model is contrasted with the output of the arbitrarily assigned process. The kappa number value ranged from 0 to 1. The value near 1 indicates the predicted model effect, while 0 implies null.

$$K = \frac{[P(a) - P(e)]}{[1 - P(e)]},$$
$$P(a) = \frac{\text{TP} + \text{TN}}{\text{N}}. \tag{20}$$

In comparison, CNN+KHHSOA remains first, and it is an emerging technique for evolving sentiment opinion analysis applications. Figure 6 shows the foremost accuracy of 96%.

Figure 7 shows the *F*-measure (%) attained by the proposed CNN+KHHSOA-based sentiment analysis method while using the dataset to distinguish opinions from facts. Sentistrength and OpinionFinder are weak in *F*-measure, being approximately 45–48 percent. Among other classifiers, the proposed technique will provide considerable results.

The predictive kappa statistic was approximately 99.8%, which confirms the suggested model is steady and adequate. From the result obtained, it was revealed that the suggested methodology expresses satisfied result when compared to other existing mechanisms (Figure 8).

Figure 9 shows the proposed tumor detection and classification method, showing a maximum precision yield of 91%, which is better than conventional methods.

Figure 10 graph shows the proposed system which shows a higher recall rate (96.5%) when compared with the system already in use. From the result obtained, it should be revealed that the suggested methodology outperforms well than any other existing methodologies.

6. Conclusion

According to the related works, classification of sentiments based on polarity is considered challenging. To overcome this challenge, in this paper, we used CNN+KHHSOA. This method uses big data to assess public opinion. A realistic and quantifiable index system is required to develop a large data multimedia network public opinion evaluation and prediction model. The new index system corrects earlier index system flaws in public opinion topics and evolutionary communication. The performance analysis of the proposed methods is discussed. The sentimental analysis gets resolute. Here, accuracy, error rate, precision, recall, and macro-F1gets were evaluated using CNN+KHHSOA. This gives the perfect results for the suggested dataset for positive and negative opinion detection. This gives the perfect results for image-text caption sentiment analysis applications.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] T. Chen, J. Rong, L. Peng, J. Yang, G. Cong, and J. Fang, "Analysis of social effects on employment promotion policies for college graduates based on data mining for online use review

in China during the COVID-19 pandemic," *Healthcare*, vol. 9, no. 7, p. 846, 2021.

[2] W. He, Y. Fang, R. Malekian, and Z. Li, "Time series analysis of online public opinions in colleges and universities and its sustainability," *Sustainability*, vol. 11, no. 13, p. 3546, 2019.

[3] H. Lee, E. B. Noh, S. H. Choi, B. Zhao, and E. W. Nam, "Determining public opinion of the COVID-19 pandemic in South Korea and Japan: social network mining on twitter," *Healthcare Informatics Research*, vol. 26, no. 4, pp. 335–343, 2020.

[4] K.-F. Wu, M. C. Chen, and B. C. Shia, "Using data mining methods to explore the important factors of university management from the perspective of school affairs research," *American Journal of Industrial and Business Management*, vol. 10, no. 8, pp. 1473–1481, 2020.

[5] J. Tang, J. Wang, and F. Li, "Research progress of network public opinion based on fuzzy set from the perspective of big data," *Journal of Physics: Conference Series*, vol. 1631, no. 1, article 012108, 2020.

[6] E. Chukwusa, H. Johnson, and W. Gao, *An exploratory analysis of public opinion and sentiments towards COVID-19 pandemic using Twitter data*, 2020.

[7] X. Lan and Z. Xu, "Scientometric analysis of research on network public opinion in a context of big data," in *2020 International Conference on Big Data and Social Sciences (ICBDSS)*, pp. 194–197, Xi'an, China, August 2020.

[8] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decision Support Systems*, vol. 57, pp. 245–257, 2014.

[9] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.

[10] F. F. Balahadia, M. C. G. Fernando, and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 95–98, Bali, Indonesia, May 2016.

[11] S. Rani and P. Kumar, "A sentiment analysis system to improve teaching and learning," *Computer*, vol. 50, no. 5, pp. 36–43, 2017.

[12] M. Misuraca, G. Scepi, and M. Spano, "Using opinion mining as an educational analytic: an integrated strategy for the analysis of students' feedback," *Studies in Educational Evaluation*, vol. 68, article 100979, 2021.

[13] M. L. Barrón Estrada, R. Zatarain Cabada, R. Oramas Bustillos, and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Systems with Applications*, vol. 150, article 113265, 2020.

[14] D. Selvapandian, K. S. Babu, R. Dhanapal, and D. Jebakumar Immanuel, "An efficient sentiment analysis on feedback assessment from student to provide better education," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 1293–1300, Palladam, India, October 2020.

[15] O. Abiodun Ayeni, A. Mercy, A. F. Thompson, and A. S. Mogaji, "Web-based student opinion mining system using sentiment analysis," *International Journal of Information Engineering & Electronic Business*, vol. 12, no. 5, pp. 33–46, 2020.

[16] A. Onan, "Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 572–589, 2021.

[17] N. T. P. Giang, T. T. Dien, and T. T. M. Khoa, "Sentiment analysis for university students' feedback," in *Future of Information and Communication Conference*, pp. 55–66, San Francisco, CA, USA, 2020.

[18] N. Nikolić, O. Grljević, and A. Kovačević, "Aspect-based sentiment analysis of reviews in the domain of higher education," *The Electronic Library*, vol. 38, no. 1, pp. 44–64, 2020.

[19] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: a systematic mapping study," *Applied Sciences*, vol. 11, no. 9, p. 3986, 2021.

[20] M. A. Toçoğlu and A. Onan, "Sentiment analysis on students' evaluation of higher educational institutions," in *International Conference on Intelligent and Fuzzy Systems*, pp. 1693–1700, Springer, 2020.

[21] Y. Yang, "Research and realization of internet public opinion analysis based on improved TF - IDF algorithm," in *2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*, pp. 80–83, Anyang, China, October 2017.

[22] J. V. Tembhurne and T. Diwan, "Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6871–6910, 2021.

[23] M. Sun and H. Zheng, "Topic detection for post bar based on LDA model," in *International conference of pioneering computer scientists, engineers and educators*, pp. 136–149, Zhengzhou, China, 2018.

[24] A. Qazi, J. Qazi, K. Naseer et al., "Adaption of distance learning to continue the academic year amid COVID-19 lockdown," *Children and Youth Services Review*, vol. 126, article 106038, 2021.

[25] K. Misiejuk, B. Wasson, and K. Egelandsdal, "Using learning analytics to understand student perceptions of peer feedback," *Computers in Human Behavior*, vol. 117, article 106658, 2021.