

IDENTIFYING SUICIDAL IDEATION: A DETECTION & PREVENTION SYSTEM

Saatvik Shukla
Department of Computer Technology
SRM Institute of Science & Technology
Kattankulathur, TN, IN
ss8639@srmist.edu.in

Dhruv Kant Ladia
Department of Computer Technology
SRM Institute of Science & Technology
Kattankulathur, TN, IN
dl6345@srmist.edu.in

C. Pretty Diana Cyril
Department of Computer Technology
SRM Institute of Science &
Technology, Kattankulathur, TN, IN
prettydc@srmist.edu.in

ABSTRACT

Social networking sites and online networks have gained immense popularity recently, providing users with a medium to share their thoughts and feelings. Regrettably, individuals increasingly hesitate to discuss mental health concerns with their friends and family. One of the major concerns today is the increasing number of suicides. Individuals may utilize social media platforms as a means to articulate their suicidal ideations. It is considered a crucial medium for self-expression, allowing users to be closely monitored and their preferences to be recorded. A recent study revealed that approximately 300 million people worldwide will be affected by depression in 2022. With the assistance of advanced technology, our goal is to detect various forms of statements that could potentially indicate suicidal ideation. Using a chatbot, we will investigate suicidal ideation behavioural patterns, which are significant

societal concerns, by utilising various Natural Language Processing and Machine Learning models. We will also develop a prevention system to prevent such incidents.

Index Terms- Depression Detection, Machine Learning Algorithms, Suicidal Ideation, Suicide Prevention

I. INTRODUCTION

Suicidal ideation refers to the act of thinking about, planning, or attempting suicide and can be triggered by a personal tragedy or a mental disorder. Active suicidal ideation specifically involves persistent and ongoing contemplation of suicide, accompanied by formulating plans and strategies to execute the act. In contrast, passive suicidal ideation only expresses a desire to die without the eagerness to commit suicide. According to data from 2022, there were 175,046 reported suicides nationwide, about a 7% increase from the previous year. As

per the World Health Organization's report, nearly 800,000 individuals die by suicide each year. While not everyone who thinks about suicide dies by suicide, it can be seen as a risk factor, underscoring the importance of timely identification and treatment of depression. Unfortunately, due to social stigma, many people experiencing depression do not seek professional help. Hence, a few individuals are turning to social media platforms to seek support. Research suggests a direct correlation between a person's mental health and their language usage. Reddit is one of the most popular social media platforms for sentiment analysis, which uses machine learning algorithms and natural language processing approaches to analyze language patterns.

II. LITERATURE REVIEW

Social networking sites and online networks have become increasingly popular as platforms for users to share their thoughts and feelings. Unfortunately, one of the biggest concerns today is the rise in suicides, with people often expressing their suicidal thoughts through social media platforms such as Twitter. Social media is recognised as an essential medium for self-expression, allowing users to be closely monitored and their preferences and interests recorded. In 2022, a recent study found that approximately 300 million individuals worldwide experienced depression. There are various types of statements that may indicate suicidal thoughts. Using advanced technologies such as Natural Language Processing(NLP) and

Machine Learning(ML) Models, we can investigate patterns of suicidal ideation and develop prevention systems to prevent such tragedies from occurring. While most studies focus on using a single machine learning system to identify despair, researchers are currently exploring various algorithms to determine which can most accurately address this issue. A chatbot is a program that indulges in multiple interactions simultaneously. It works like the human brain using artificial intelligence, machine learning and neural capabilities. Chatbots speak as easily as humans. Chatbots caught his attention in 2016. 2016 is said to be the first year of chatbots. More than many startups and companies have started using chatbots to improve customer service. Research shows that chatbots are now being used in railway bookings, bus bookings, stay bookings, logistics, and businesses like Amazon and Flipkart. Analyse common queries based on patterns. Artificial Intelligence uses AI/ML in particular. The main advantage of using his AI/ML in chatbots is that it is easy to use and learn. The salient characteristics of chatbots are their user-friendly interface, conversational system, and capacity to operate in multiple languages, with the ability to create structured, computer-readable representations. Many startups have developed numerous chatbots, which organizations have employed to enhance their customer service and provide prompt, compassionate responses. Studies indicate that presently, chatbots are employed in diverse industries, such as e-commerce, insurance, banking, healthcare, finance, legal,

telecommunications, logistics, retail, automotive, leisure, travel, sports, entertainment, media, and more. Various organizations are now using chatbots to respond quickly and productively, and some even post occasional customer inquiries. [Reference number-1]

The utilization of therapy chatbots has the potential to revolutionize the mental health sector and provide numerous advantages to both patients and healthcare practitioners. They can improve the efficiency of mental health services by automating administrative tasks, allowing practitioners to focus on delivering personalized patient care. This can help combat burnout among healthcare professionals and improve the quality of treatments offered to mental health patients. Moreover, therapy chatbots can offer accessible and affordable healthcare support to thousands of patients who may not have access to mental health services due to several reasons. They can also provide patients with doctor-approved information and resources to self-diagnose their symptoms, thereby enabling them to manage their symptoms and seek appropriate help. Additionally, therapy chatbots can monitor patient-reported outcomes and provide insights that can help healthcare professionals better diagnose and treat mental health patients. They can also be used to set up online appointments, thereby eliminating the need for patients to travel to hospitals and clinics to seek help. However, it is essential to note that further research is needed before therapy chatbots can

be widely adopted. The chatbots must be designed to protect the personal health information of patients, and the technology must be continually monitored and updated to ensure that it remains effective and accurate. There are four main advantages to mental healthcare chatbots:

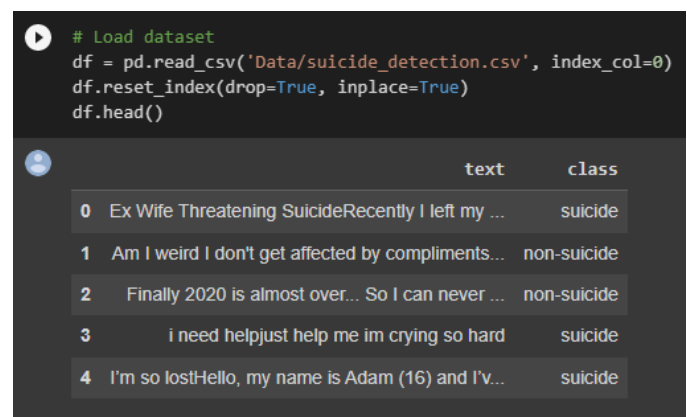
- Continuous availability
- Instantly accessible in critical situations
- Data collection through patient engagement
- Handling high patient volumes with ease

[Reference number-12].

III. PROPOSED WORK

A. Dataset Collection:

The dataset utilized for detecting suicidal ideation in this project has been acquired from Kaggle. It consists of textual posts from the social media platform, Reddit.



```
# Load dataset
df = pd.read_csv('Data/suicide_detection.csv', index_col=0)
df.reset_index(drop=True, inplace=True)
df.head()
```

	text	class
0	Ex Wife Threatening SuicideRecently I left my ...	suicide
1	Am I weird I don't get affected by compliments...	non-suicide
2	Finally 2020 is almost over... So I can never ...	non-suicide
3	i need helpjust help me im crying so hard	suicide
4	I'm so lostHello, my name is Adam (16) and I've...	suicide

Figure 1: A Fraction of Dataset on Display

The dataset has been labelled using a binary classification approach, where posts from subreddits 'SuicideWatch' are labelled "suicidal" and posts from 'teenagers' are labelled "non-suicidal". It's important to note that this type of labelling approach may have some limitations, as there may be some posts that do not fit into either of these categories. Nonetheless, it is a good starting point for building a model to identify suicidal ideation in social media posts.

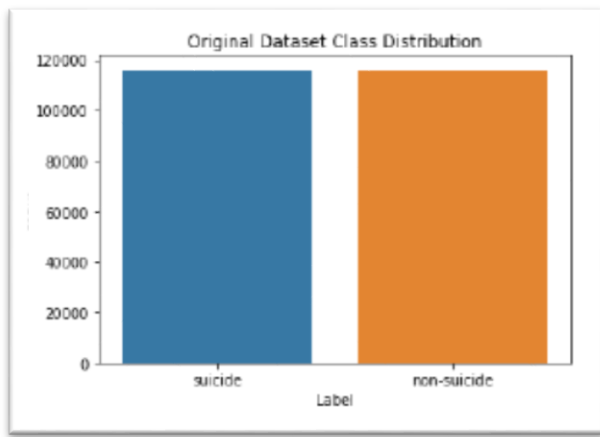


Figure 2: Original Dataset Class Distribution

The classes are equally distributed, as seen in the dataset where there are 116,037 rows, representing 50% of the dataset, within each class.

B. Text Pre-Processing:

Prior to model building, the text data necessitates pre-processing to transform it into suitable formats. Social media data, specifically, is typically less structured and requires more tailored preprocessing and cleansing techniques. Thus, our data was cleaned with the following steps in the sequence seen in Figure below.

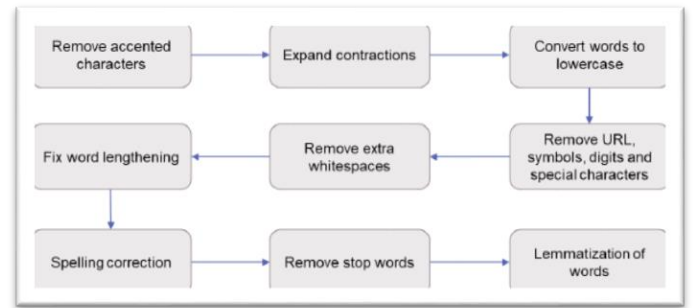


Figure 3: Text Pre-Processing in Detail

Text Pre-processing is carried out to remove accented characters, expand contractions, convert to lowercase, remove website url(s), symbols and digits, special characters, and extra whitespaces, work on word lengthening, spelling correction, remove stop-words, lemmatization.

C. Data Cleaning:

Removing irrelevant words and Outlier Rows with High Word Count.

D. MODEL BUILDING & EVALUATION

We will be building different models and evaluating their performance in classifying suicidal text. The objective of our problem statement is to make predictions on a binary variable that distinguishes between suicide and non-suicide.

Logistic Regression (Logit) is a simple yet powerful statistical model used for binary classification. In this model, the probability of a positive class is estimated using a logistic function of the input features. The Convolutional

Neural Network (CNN) is a type of deep learning model that is frequently utilized for image and text classification. CNNs use convolutional layers to extract significant features from input data and a fully connected layer to forecast the output class. In contrast, Long Short-term Memory (LSTM) is a form of recurrent neural network (RNN) created to manage sequential data by retaining long-term dependencies. LSTMs use a combination of memory cells and gates to selectively forget and remember input data. Efficiently Learning on Encoder that Classifies Token Replacements Accurately (ELECTRA) is a transformer-based language model that employs a pre-training task known as replaced token detection to acquire high-quality representations of natural language text. ELECTRA has demonstrated excellent outcomes in several natural language processing (NLP) tasks, including text classification.

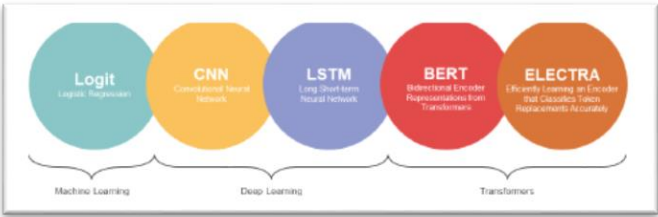


Figure 4: Important Machine Learning Models

E. Model Selection:

The transformer model, ELECTRA, has performed exceptionally well in your analysis. Transformers have been known to outperform traditional machine learning models on various natural language processing tasks. It's also

great to see that the customized Word2Vec embeddings have improved the performance of the other models.

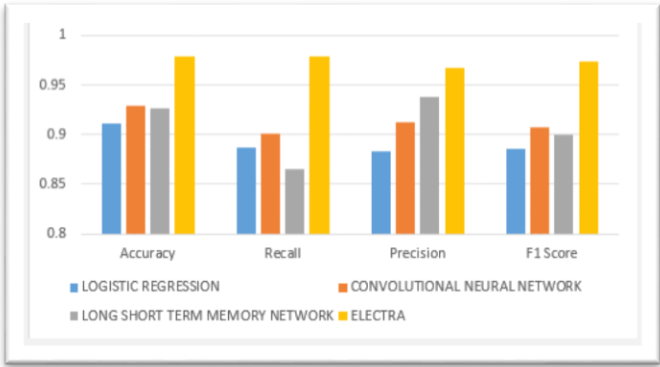


Figure 5: Models' Score Graph

IV. IMPLEMENTATION

A. ELECTRA:

ELECTRA has been shown to outperform other pre-trained transformer models like BERT, RoBERTa, and XLNet on several benchmark datasets while requiring less computational power. ELECTRA's superior performance can be attributed to its implementation of the Replaced Token Detection (RTD) pre-training task, as opposed to BERT's use of the Masked Language Model (MLM) task. RTD allows ELECTRA to train on all input tokens rather than just the masked ones, resulting in a more efficient use of training data.

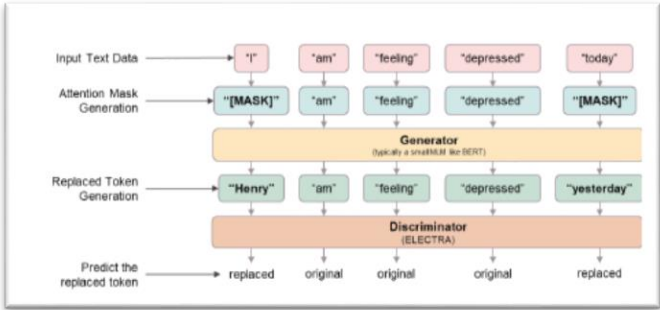


Figure 6: Replaced Token Detection (RTD)

ELECTRA's RTD task works by creating two versions of the input data: the original data and the corrupted data. The corrupted data is generated by replacing some of the tokens in the original data with other tokens. The model is then trained to distinguish between the original and corrupted data, which helps it to learn the relationship between the input tokens and their context more effectively. Compared to BERT's MLM technique, ELECTRA's RTD task allows the model to learn from every single input token, not just the masked tokens, which results in a more efficient and effective pre-training process. ELECTRA also introduces a more efficient generator network that creates the corrupted data, which contributes to the model's improved performance while requiring less computational power. In contrast, ELECTRA uses a different approach called RTD to train a bidirectional model. In this method, tokens are replaced with incorrect but plausible alternatives instead of the usual "[MASK]" tokens used in BERT. For example, the sentence "I am feeling depressed today" may be transformed into "Henry am feeling depressed yesterday". Although the replaced sentence makes more sense than using "[MASK]" tokens, it still does not perfectly fit the context.

The discriminator in ELECTRA is then trained to identify the replaced tokens, similar to how generative adversarial networks (GANs) distinguish between real and fake input data. This binary classification task promotes more accurate representation learning, as the model

must learn an accurate data representation to solve the task. Compared to BERT, ELECTRA can achieve better performance with lesser training data due to its ability to extract more information per example.

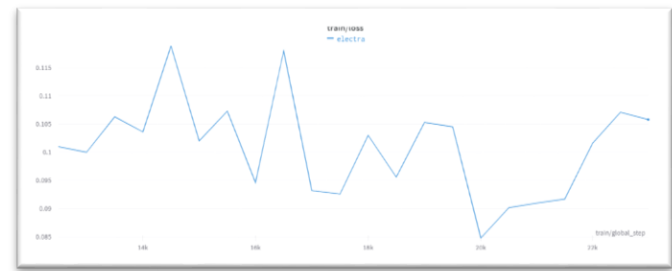


Figure 7: Train/Loss Graph

Below is a summary of the model performance for all ELECTRA model variants. Model 2, which is the fine-tuned ELECTRA, performed significantly better than Model 1, the pre-trained ELECTRA, on accuracy, precision, and F1 score. While Model 1 had a higher recall score, which was close to 1, this did not affect the model choice since the focus was placed on the F1 score.

In summary, Model 2 was the best-performing variant, while Model 1 had a high recall score but lower overall performance compared to Model 2.

ELECTRA Model Variants	Accuracy	Recall	Precision	F1 Score
1. Pre-trained ELECTRA	0.4025	0.9908	0.3918	0.5615
2. Fine-tuned ELECTRA	0.9792	0.9788	0.9677	0.9732

Figure 8: ELECTRA Models' scores

B. Chatbot Integration:

Our chatbot utilises transformers, which have a self-attention mechanism that can enhance performance compared to traditional recurrent

methods. The generative aspect of the chatbot is powered by Dialo-GPT, which is a pre-trained transformer-based model designed for multi-turn conversations and was created by Microsoft in 2019. Dialo-GPT was built on 147 million multi-turn dialogue from Reddit discussion threads and is available in three variations, namely small, medium, and large, similar to other transformer models. We utilised the original pre-trained Dialo-GPT model from the Hugging Face Transformers library. Although we attempted to train a customised chatbot, we needed a large chat dataset and training resources to generate coherent and relevant responses. Nevertheless, according to a single-turn conversation Turing test conducted by Zhang et al. (2019), responses generated by Dialo-GPT were comparable to human response quality. It should be noted that the pre-trained Dialo-GPT chatbot is not equipped to provide appropriate replies to suicidal messages. We added a retrieval-based component to the chatbot that matches users' input with a library of comforting messages from various suicidal prevention websites and a list of local helplines for immediate support. Whenever a user input suggests suicidal intent, the chatbot will respond with the appropriate message to offer comfort and helpful information.

Neural Response Generation (NRG) Model

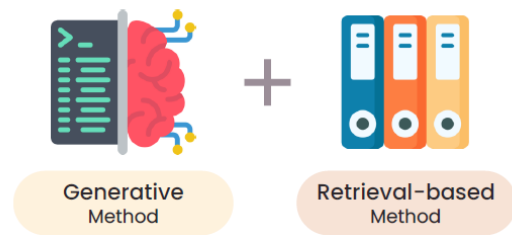


Figure 9: Chatbot Architecture

C. LIMITATIONS:

I agree that chatbots can be a valuable tool in the mental health, particularly in providing support and resources to individuals who may not have access to traditional mental health services. However, as you mentioned, there are limitations to chatbots' ability to provide empathetic responses and accurately identify suicidal risks. Therefore, it is important to continue researching and developing chatbots that can provide more nuanced and effective answers. Additionally, ensuring the privacy and security of personal health information is crucial for the widespread adoption of mental health chatbots. Overall, chatbots have the potential to be a valuable complement to traditional mental health services, but careful consideration and development are necessary to ensure their effectiveness and ethical use.

RESULTS AND DISCUSSION

Our project has accomplished the goal of identifying suicidal text in social media posts.

Our models have produced outstanding results, with ELECTRA achieving the highest F1 score of 0.973. Furthermore, we have integrated our detection model into a fully operational chatbot, allowing us to support people who require assistance. As we advance, we aim to improve the performance of our detection model and chatbot while also broadening our scope to assist more needy individuals.

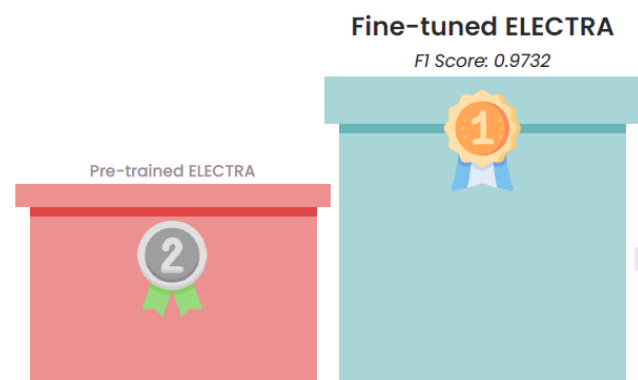


Figure 10: ELECTRA valuation

```

start_chatbot()

==== Hello! I am Mikasa and I am your virtual friend. If you need a listening ear, I'm a
>> You: hi
A decoder-only architecture is being used, but right-padding was detected! For correct g
Mikasa: hi
>> You: i am feeling like crying
A decoder-only architecture is being used, but right-padding was detected! For correct g
Mikasa: i am crying
>> You: what do i do
A decoder-only architecture is being used, but right-padding was detected! For correct g
Mikasa: i am crying
>> You: i want to kill myself
A decoder-only architecture is being used, but right-padding was detected! For correct g
Mikasa: Is there anything I can do to make this day go easier for you?
In times of severe distress where you need to speak with someone immediately, these are suicide hotline serv
volunteers or professionals who are trained to deal with suicide crisis. Mental Health Rehabilitation Helpline KI
>> You: exit
Alex: See ya
  
```

Figure 11: Final Chatbot Implementation

CONCLUSION

Through ELECTRA transformer technology, we have likely created a chatbot capable of engaging in natural and fluid conversations with

users while providing valuable mental health support and resources. Overall, our research highlights the potential of modern machine learning techniques to support mental healthcare and improve people's quality of life. By leveraging the power of ELECTRA transformer technology, you have demonstrated how advanced natural language processing capabilities can be used to develop innovative and effective mental health tools. It is hoped that our work will inspire further research in this area and contribute to the development of more sophisticated mental healthcare chatbots in the future and support individuals who may be struggling with mental health issues. Addressing mental health issues and providing timely support to struggling people is essential.

REFERENCES

1. Arya, Vanshika & Khan, Rukhsar & Aggarwal, Prof. (2022). A Chatbot Application by using Natural Language Processing and Artificial Intelligence Markup Language. International Journal of Soft Computing and Engineering. 12. 1-7. 10.35940/ijscce.C3566.0712322.
2. Tirumala, Kushal & Markosyan, Aram & Zettlemoyer, Luke & Aghajanyan, Armen. (2022). Memorization Without Overfitting: Analyzing the Training Dynamics of Large Language Models. 10.48550/arXiv.2205.10770.
3. Zhang, Xiaopeng & Qin, Liangxi. (2022). An Improved Extreme Learning Machine for Imbalanced Data Classification. IEEE

- Access. PP. 1-1.
10.1109/ACCESS.2022.3142724.
4. Rahali, Abir & Akhloufi, Moulay. (2023). End-to-End Transformer-Based Models in Textual-Based NLP. AI. 4. 54-110. 10.3390/ai4010004.
 5. Gabín, Jorge & Ares, M. & Parapar, Javier. (2023). Keyword Embeddings for Query Suggestion. 10.48550/arXiv.2301.08006.
 6. Bao, Han-Wu-Shuang & Wang, Zi-Xi & Cheng, Xi & Su, Zhan & Yang, Ying & Zhang, Guang-Yao & Wang, Bo & Cai, Huajian. (2023). Using word embeddings to investigate human psychology: Methods and applications. 31. 887. 10.3724/SP.J.1042.2023.00887.
 7. Ricciardelli, Elena & Biswas, Debmalya. (2019). Self-improving Chatbots based on Reinforcement Learning.
 8. Cameron, Gillian & Cameron, David & Megaw, Gavin & Bond, Raymond & Mulvenna, Maurice & O' Neill, Siobhan & Armour, Cherie & Mctear, Michael. (2019). Assessing the Usability of a Chatbot for Mental Health Care. 10.1007/978-3-030-17705-8_11.
 9. Ce, Peng & Tie, Bao. (2020). An Analysis Method for Interpretability of CNN Text Classification Model. Future Internet. 12. 228. 10.3390/fi12120228.
 10. Barbosa Pina, Débora & Kunstmann, Liliane & Bevilaqua, Felipe & Siqueira, Isabela & Lyra, Alan & de Oliveira, Daniel & Mattoso, Marta. (2022). Capturing Provenance from Deep Learning Applications Using Keras-Prov and Colab: a Practical Approach. Journal of Information and Data Management. 13. 10.5753/jidm.2022.2544.
 11. Ni, Shiwen & Kao, Hung-Yu. (2022). ELECTRA is a Zero-Shot Learner, Too. 10.48550/arXiv.2207.08141.
 12. Viduani, Anna & Cosenza, Victor & Araújo, Ricardo & Kieling, Christian. (2023). Chatbots in the Field of Mental Health: Challenges and Opportunities. 10.1007/978-3-031-10698-9_8.

Major Project-Identifying Suicidal Ideation: A Detection & Prevention System

ORIGINALITY REPORT

9%

SIMILARITY INDEX

4%

INTERNET SOURCES

3%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1

ijrar.org

Internet Source

1%

2

Submitted to Birla Institute of Technology and Science Pilani

Student Paper

<1%

3

Submitted to International University - VNUHCM

Student Paper

<1%

4

Submitted to University of Portsmouth

Student Paper

<1%

5

Submitted to University of Ulster

Student Paper

<1%

6

academic-accelerator.com

Internet Source

<1%

7

"Neural Information Processing", Springer Science and Business Media LLC, 2017

Publication

<1%

8

www.mdpi.com

Internet Source

<1%

9	link.springer.com Internet Source	<1 %
10	Submitted to University of Bedfordshire Student Paper	<1 %
11	Submitted to University of East London Student Paper	<1 %
12	Submitted to University of North Carolina, Greensboro Student Paper	<1 %
13	Biniyam Mulugeta Abuhayi, Abdela Ahmed Musa. "Coffee disease classification using Convolutional Neural Network based on feature concatenation", Informatics in Medicine Unlocked, 2023 Publication	<1 %
14	Submitted to Sheffield Hallam University Student Paper	<1 %
15	Submitted to Symbiosis International University Student Paper	<1 %
16	Submitted to Florida Atlantic University Student Paper	<1 %
17	Submitted to Middlesex University Student Paper	<1 %
18	bagotayo.net Internet Source	<1 %

19	chatbotsjournal.com Internet Source	<1 %
20	Submitted to Al Akhawayn University in Ifrane Student Paper	<1 %
21	Submitted to Istanbul Aydin University Student Paper	<1 %
22	Submitted to Liverpool John Moores University Student Paper	<1 %
23	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
24	forgedigitalmarketing.com Internet Source	<1 %
25	Anfu Guo, Dekun Kong, Xiaoyan Zhou, He Kong, Peng Qu, Shaoqing Wang, Hongbing Wang, Yingbin Hu. "Method for Preparing Damage-Resistant 3D-Printed Ceramics via Interior-to-Exterior Strengthening and Toughening", Additive Manufacturing, 2022 Publication	<1 %
26	Submitted to California Southern University Student Paper	<1 %
27	Submitted to Fr Gabriel Richard High School Student Paper	<1 %

28	Submitted to Herzing University Student Paper	<1 %
29	download.atlantis-press.com Internet Source	<1 %
30	www.techscience.com Internet Source	<1 %
31	Submitted to Anatolia College Student Paper	<1 %
32	Submitted to Luton Sixth Form College, Bedfordshire Student Paper	<1 %
33	www.fortinet.com Internet Source	<1 %
34	Submitted to Carnegie Mellon University Student Paper	<1 %
35	Sangyoup Lee, Eunsu Lee, Menachem Elimelech, Seungkwan Hong. "Membrane characterization by dynamic hysteresis: Measurements, mechanisms, and implications for membrane fouling", Journal of Membrane Science, 2011 Publication	<1 %
36	scindeks-clanci.ceon.rs Internet Source	<1 %
37	www.disabilityrightsca.org Internet Source	

<1 %

38

www.tandfonline.com

Internet Source

<1 %

39

"Information, Communication and Computing Technology", Springer Science and Business Media LLC, 2019

Publication

<1 %

40

Lizon Maharjan, Mark Ditsworth, Babak Fahimi. "Critical Reliability Improvement Using Q-Learning-Based Energy Management System for Microgrids", Energies, 2022

Publication

<1 %

41

Tariq J. Al-Musawi, Narjes Sadat Mazari Moghaddam, Seyedeh Masoomah Rahimi, Matin Hajjizadeh, Negin Nasseh. "Hexadecyltrimethylammonium-activated and zinc oxide-coated nano-bentonite: A promising photocatalyst for tetracycline degradation", Sustainable Energy Technologies and Assessments, 2022

Publication

<1 %

42

Submitted to University of Westminster

Student Paper

<1 %

43

ijeecs.iaescore.com

Internet Source

<1 %

44	medinform.jmir.org Internet Source	<1 %
45	tel.archives-ouvertes.fr Internet Source	<1 %
46	www.telegraphindia.com Internet Source	<1 %
47	Andika Lin, Nicholas Livando, William Chandra, Gary Phan, Amir Mahmud Husein. "Sentiment Analysis Of Hotel Reviews On Tripadvisor With LSTM And ELECTRA", SinkrOn, 2023 Publication	<1 %
48	Liset Vázquez Romaguera, Rosalie Plantefève, Francisco Perdigón Romero, François Hébert et al. "Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks", Medical Image Analysis, 2020 Publication	<1 %
49	hdl.handle.net Internet Source	<1 %
50	Submitted to University of Queensland Student Paper	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On